DISSERTATION

THE CURIOUS CASE OF CHEMOTAXIS IN SOFT ROT PECTOBACTERIACEAE

Submitted by

Afnan Shazwan Nasaruddin

Department of Agricultural Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2021

Doctoral Committee:

Advisor: Amy O. Charkowski

Jan E. Leach
Pankaj Trivedi
Adam L. Heuberger

ABSTRACT

THE CURIOUS CASE OF CHEMOTAXIS IN SOFT ROT PECTOBACTERIACEAE

Soft rot Pectobacteriaceae, *Dickeya* and *Pectobacterium,* are notorious for causing blackleg and soft rot diseases on more than 50% angiosperms such as potato, tomato, carrot, cabbage, and rice. In the United States, soft rot Pectobacteriaceae causes at least $40 million losses of potato each year. Flagellar motility is important for soft rot Pectobacteriaceae virulence. Chemotaxis, which controls flagellar motility towards a conducive environment or away from hostile conditions, is essential for initial stages of infection. Chemotaxis is mediated by chemoreceptors known as methyl-accepting chemotaxis proteins.

Genomic analyses of soft rot Pectobacteriaceae and closely related animal pathogens and non-pathogenic bacteria in order Enterobacteriales showed that soft rot Pectobacteriaceae genomes are enriched in methyl-accepting chemotaxis proteins. Furthermore, soft rot Pectobacteriaceae methyl-accepting chemotaxis proteins contain more diverse ligand binding domains compared to other species in Enterobacteriales. This study suggests the importance of chemotaxis for soft rot Pectobacteriaceae pathogenicity and opens up possibilities for future research in targeting chemotaxis for plant disease management.

In *E. coli,* the alternative sigma factor FliA is required for transcription initiation in motility and chemotaxis genes. To determine how chemotaxis is regulated in *Dickeya*, we conducted an RNA-sequencing experiment using a wild-type strain and a *fliA* mutant of *D. dadantii* 3937 grown in minimal media with glycerol or glucose. We found that the FliA sigma factor did not regulate methyl-accepting chemotaxis genes in *Dickeya,* several virulence genes were upregulated in glucose, and some genes postulated to be regulated by PecS were

upregulated in glycerol. It is still a mystery as to which sigma factor regulates the chemotaxis genes in *Dickeya*, however, my work demonstrates that the regulation of chemotaxis in plant pathogens differs from closely related animal pathogens in the same order.

ACKNOWLEDGEMENTS

This dissertation would never have been possible without the contributions of countless people throughout these past five years and a half. First and foremost, I would like to express my gratitude to Amy, for being the best mentor that I could ever ask for. Thank you for your patience in guiding me throughout these years, for believing in my potential, for showing me my strengths, for helping me deal with my depression and imposter syndrome, for lending your ears to my problems and offering invaluable advices, and for your kindness and tough love. I can only hope to have lived up to your expectations.

I would also like to thank each member of my committee for their contributions in my scientific training. Thank you for validating my intended career path in teaching. Your support means the world to me. None of the work could have been done without either the funding provided by College of Agricultural Sciences at Colorado State University, USDA-ARS, APHIS, and Wisconsin Potato and Vegetable Association, or the administrative support from the Agricultural Biology department and Agricultural Business Center offices. Thank you all.

I must also thank Ana Cristina Fulladolsa Palma, Alejandra Huerta, and Yuan Zeng for their mentoring, friendship, and encouragement over the years. I will truly miss all the times we have spent together. Thank you for always being there for me.

Thank you to the Charkowski lab members and fellow graduate students at University of Wisconsin-Madison and Colorado State University, both past and present, for sharing this experience with me and providing a lending hand or a listening ear in times of need.

Last but not least, I would like to thank my family for their love, support, and patience.

PREFACE

Chapter 1 provides background to the research and contains preliminary results that led to the studies in Chapter 2 and 3.

Chapter 2 will be submitted to *BMC Genomics* as:

> Nasaruddin AS, Zeng Y, Yedidia I, and Charkowski AO. "Comparative genomic analyses of the order Enterobacteriales support the importance of ATP-binding cassette transporters and methyl-accepting chemotaxis proteins in the pathogenicity of soft rot Pectobacteriaceae."

Both ASN and YZ contributed equally to this chapter. ASN contribution involved methyl-accepting chemotaxis proteins, while YZ contributed to the ATP-binding cassette transporters.

Chapter 3 is in the process of editing and will be submitted to *Molecular Plant-Microbe Interactions* as:

> Nasaruddin AS, Zeng Y, and Charkowski AO. "Elucidation of FliA and glucose regulation in *Dickeya dadantii* 3937 in minimal media."

DEDICATION

To anyone who has ever doubted their potential, you are more than enough, and you will overcome any obstacles and achieve your dream. Your dream is worth fighting for.

I dedicate this dissertation to my family, for letting me be more than 8,000 miles from them, and for supporting me all these years.

TABLE OF CONTENTS

Chapter 1

General Introduction and Preliminary Studies

BACKGROUND

Potato is the most important non-grain crop worldwide. In the United States, the production value of potato is approximately $4 billion annually. Each year, at least $40 million of potato losses are attributed to bacterial disease known as blackleg and soft rot of potato. Blackleg and soft rot disease of potato are caused by two genera of Gram-negative bacteria called *Dickeya* and *Pectobacterium* (1). These phytopathogenic bacteria belong in the family Pectobacteriaceae, together with the plant pathogens *Brenneria* and *Lonsdalea*, and the non-pathogenic *Soldalis* and *Biostraticola* (2). *Dickeya* and *Pectobacterium* are rod-shaped, facultative anaerobes that can infect a broad host range of plants such as potato, rice, chicory, carrot, tomato, cabbage, broccoli, peppers, and ornamentals (*Begonia, Dahlia, Hyacinthus,* and *Iris*) (1, 3, 4). These bacteria are ubiquitous and can be found in the soil, clouds, open water sources such as lake, crop residues and weeds (1, 3). Both *Dickeya* and *Pectobacterium* infect their host plants through natural openings such as stomata, or mechanical wounds (1, 3). They can be spread through irrigation water, infected seed potatoes, farm equipment and tools, and insects (1, 3).

The key virulence factor of soft-rot Pectobacteriaceae (SRP) is the various plant cell wall degrading enzymes (PCWDEs) secreted through the Out type II secretion system (T2SS) such as pectate lyases, cellulases, and pectinases, which macerate the plant tissues, causing the characteristic blackleg and soft-rotting symptoms (1, 5). Unlike the model plant pathogenic

bacterium *Pseudomonas syringae* that has numerous effector proteins secreted via the type III secretion system (T3SS), SRP only have one known T3-effector protein, DspA/E (5, 6). Compared to the homologs present in other phytopathogens such as *Erwinia amylovora*, the DspE allele in SRP is truncated, and the effector was unable to prevent callose formation in leaves (6). Interestingly, the DspE of SRP can trigger programmed cell death in *Nicotiana benthamiana* leaves, causing leaf maceration, which is beneficial for the necrotrophic SRP (6).

Motility is important for the virulence of some SRP, as non-motile mutants of *Pectobacterium carotovorum* and *Dickeya dadantii* had reduced symptom development in Chinese cabbage and *N. benthamiana*, respectively, compared to the wild-type strain (7, 8). SRP have peritrichous flagella, which are secreted through the flagellar T3SS, a subtype of the T3SS (5). In contrast to *P. syringae*, SRP are motile during infection (9). SRP genomes encode flagellar genes that are homologous to other bacteria in the order Enterobacteriales, such as *Escherichia coli* (8). Flagella enable bacteria to actively explore their surroundings and swim towards favorable conditions or away from hostile environments using chemotaxis, a process by which bacterial cells sense and respond to chemical signals by altering the rotation of the flagella (10).

Numerous studies have shown the pivotal role of chemotaxis for pathogenic bacteria in initial stages of infection, particularly in locating their host (11). For example, *E. coli* serotype O157:H7 that causes bloody diarrhea in humans, finds the epithelial cell surface in the gastrointestinal tract by chemotaxis towards hormones epinephrine and norepinephrine (12). The plant pathogen *D. dadantii* 3937 is chemotactically attracted to the plant hormone jasmonic acid, which is a common signaling compound in wounded plant tissues (13). Furthermore, *D. dadantii* 3937 spreads more on the leaf surface of wounded chicory leaves compared to unwounded

leaves, suggesting that chemotaxis to jasmonic acid aids in bacterial entry into the host plant via wounded tissues (13).

Chemotaxis is a complex two-component system that is mediated by chemoreceptors known as methyl-accepting chemotaxis proteins (MCP) (10, 14). In the bacterial cells, MCP are organized as trimers of dimers at the cell pole (15, 16). Typically, MCP are comprised of four domains, 1) a ligand binding domain (LBD), 2) transmembrane helix (TM), 3) a histidine kinase, adenyl cyclase, methyl-accepting chemotaxis protein and phosphatase (HAMP) domain, and 4) a cytoplasmic signaling domain (SD) (Figure 1.1) (10).

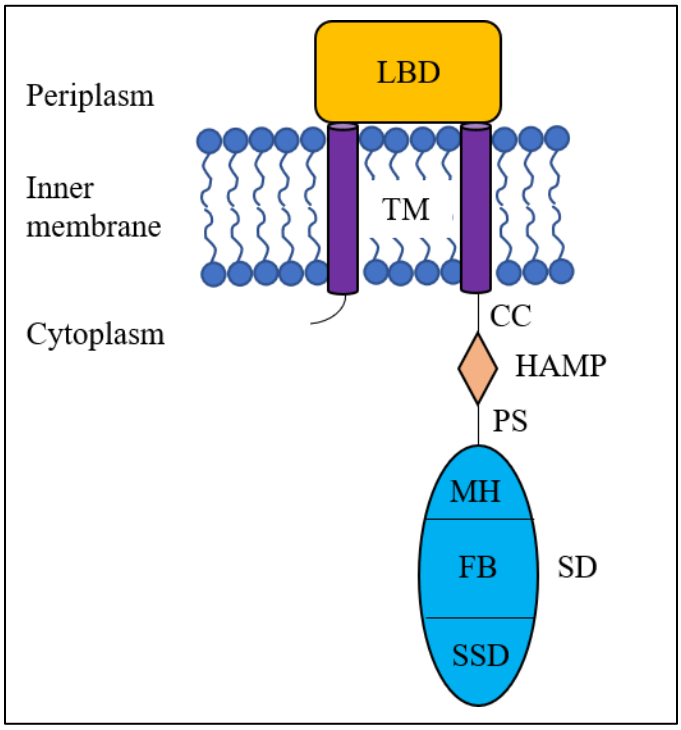

Figure 1.1: Typical topology of methyl-accepting chemotaxis protein (10). LBD: ligand binding domain; TM: transmembrane helix; CC: control cable; HAMP: histidine kinase, adenyl cyclase, methyl-accepting chemotaxis protein and phosphatase domain; PS: phase stutter; SD: signaling domain; MH: methylation helix; FB: flexible bundle; SSD: signaling subdomain.

The LBD is responsible for sensing extracellular or intracellular signals, depending on the type of the LBD (10). MCP can sense their ligands through direct binding via the LBD (17, 18) or through interaction with periplasmic binding proteins such as ATP-binding cassette

(ABC) transporters (19-22). Examples of periplasmic LBD include the four-helix bundle (4HB) domain, helical bimodular (HBM) domain, single or double CACHE (calcium channels and chemotaxis receptors) domain, and CHASE (cyclase/histidine kinase-associated sensory extracellular) domain (23-28). Known examples of cytoplasmic LBD are the PAS (found in Per-Arnt-Sim proteins) domain and the NIT (nitrate and nitrite binding) domain (23, 29, 30). The PAS domain contains bound heme, flavin adenine dinucleotide (FAD), or flavin mononucleotide (FMN) involved in aerotaxis or energy sensing through oxygen and redox sensing (23, 31). The NIT domain is involved with nitrate and nitrite sensing and is postulated to be a redox sensor since anaerobic metabolism is tightly controlled by nitrate in *E. coli* (29, 32). Interestingly, NIT domain can also be found in the periplasm (29). In addition, one LBD known as FIST (F-box intracellular signal transduction) domain can be either intracellular and extracellular depending on the presence of signal peptide (33).

Functional annotation of the types of signals LBD sense is difficult, because the LBD has the lowest sequence conservation among the four MCP domains (34). In addition, LBD of the same type can sense different signals and different types of LBD can sense the same signals (23). *E. coli* has five MCP in which four of them (Tar, Tsr, Trg, and Tap) have the 4HB domain, but each mediates response to different ligands (23). Tar binds to aspartate, Tsr recognizes serine, Trg senses ribose and galactose, and Tap is attracted to pyrimidines and dipeptides (17, 35-38). Furthermore, LBD comprised of 4HB, double CACHE, and HBM domains can sense the same signals, namely di- and tricarboxylic acids (23).

The signals sensed by the LBD are transmitted through the transmembrane helices that are connected to the HAMP domain by a five-residue control cable (CC) down to the cytoplasmic SD, which is connected to the HAMP domain via a four-residue phase stutter (PS)

(10, 39). The cytoplasmic SD consists of three subdomains, methylation helix (MH), flexible bundle (FB), and the signaling subdomain (SSD) (Figure 1.1) (10). The MH subdomain is where the signal-dependent reversible methylation of MCP occurs (10, 39). The FB subdomain contains the glycine hinge, which is presumed to be important for the formation of MCP trimers (39). The SSD is the part of MCP that directly interacts with some components of the chemotaxis system (10, 39).

Besides MCP, there are core chemotaxis (Che) proteins involved in this two-component signal transduction system, namely CheA, CheW, CheV, CheY, CheZ, CheB, and CheR (Figure 1.2) (10, 14). MCP cytoplasmic SSD interacts with the sensor-less histidine kinase CheA and the scaffolding protein CheW, forming a stable ternary signaling complex (10, 14). Besides CheW, CheV can also form a complex with MCP and CheA (40). CheA is able to autophosphorylates itself using ATP as a substrate when it is in a complex with CheW and methylated MCP (10, 14). CheY is the response regulator that when phosphorylated by phosphorylated CheA, interacts with the flagellar motor, inducing the clockwise (CW) rotation of flagella that causes a tumble (10, 14). CheZ is a phosphatase that dephosphorylates CheY, returning the flagella to the default rotation of counterclockwise (CCW) that enables bacterial cell to swim in a straight direction (run) (10, 14). Both CheB and CheR are part of the chemotaxis adaptation pathway that restores the balance between CW and CCW outputs (i.e. returning the cell back to its initial swimming pattern), through demethylation and methylation of MCP at the MH subdomain, respectively (10, 14, 41). The relative time of tumble and run and the signal extinction from the MCP result in a net movement towards favorable conditions and away from hostile environments.

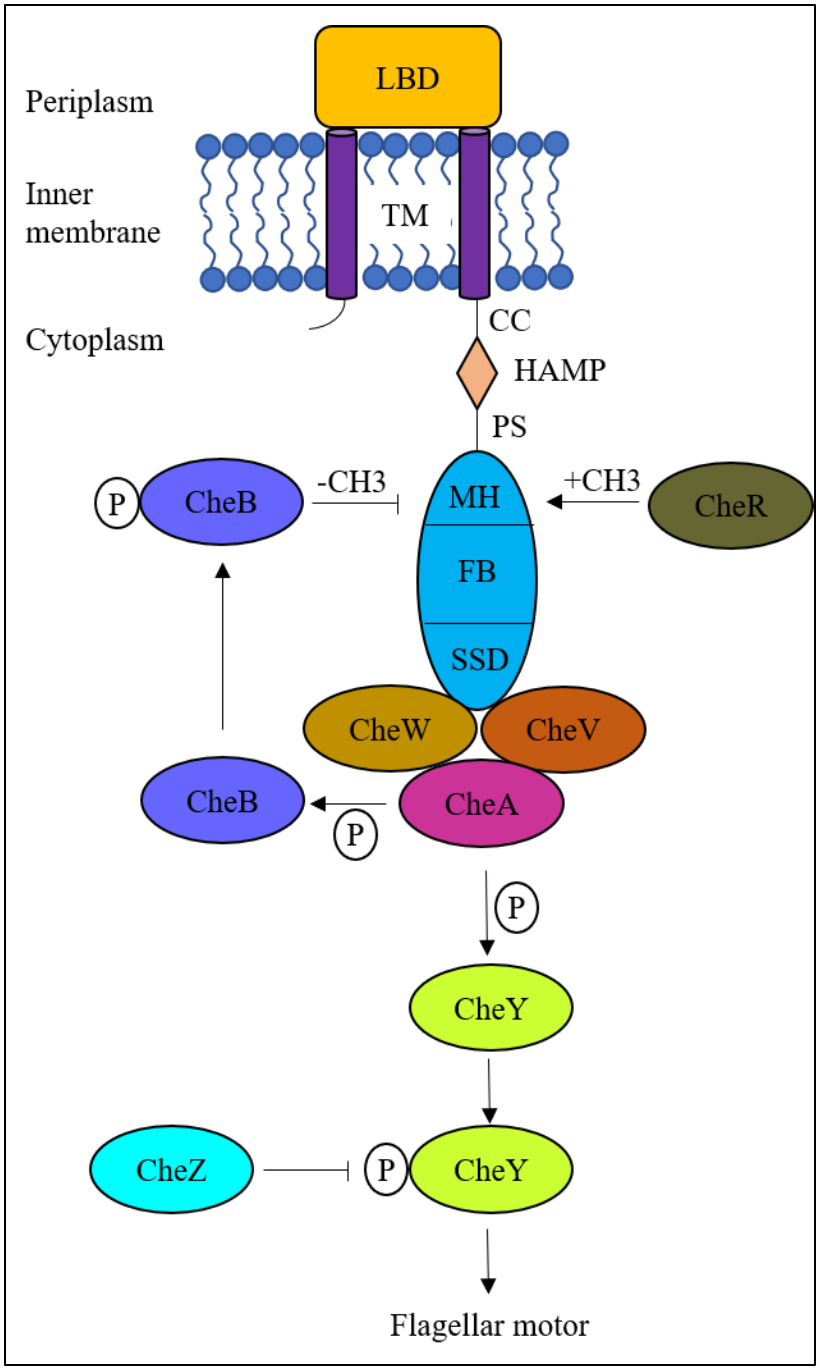

Figure 1.2: Schematic of the complex two-component signal transduction system of chemotaxis as studied in *E. coli*. (10). LBD: ligand binding domain; TM: transmembrane helix; CC: control cable; HAMP: histidine kinase, adenyl cyclase, methyl-accepting chemotaxis protein and phosphatase domain; PS: phase stutter; MH: methylation helix; FB: flexible bundle; SSD: signaling subdomain.

On average, animal pathogenic bacteria encode 17 MCP genes per genome, while plant pathogenic bacteria encode 33 MCP (24). The relatively high number of MCP genes in phytopathogenic bacteria might be attributed to their complex lifestyles and the need to adapt to multiple and fluctuating environments (24). The high number of MCP genes can be seen with plant pathogens such as *Pseudomonas syringae* and *Ralstonia solanacearum.* However, there is an exception to the rule. One notable example is *Xylella fastidiosa*, whose habitat is limited to the xylem and insect gut. The current knowledge about these three phytopathogens is described below.

***Pseudomonas syringae* pv. tomato**

Pseudomonas syringae pv. tomato is a Gram-negative, rod-shaped, aerobic, foliar bacterium that is the causal pathogen for bacterial speck of tomato (42). It can grow as a saprophyte or an epiphyte on plant debris, soil, and leaf surfaces (43). Motility and chemotaxis play an important role for this pathogen to locate and enter its host plant through stomatal opening or wounds (42). *P. syringae* pv. tomato strain DC3000 genome encodes five clusters of chemotaxis proteins and 49 MCP (43). Both cluster I and V are essential for chemotaxis since they encode the chemotaxis genes *cheA, cheB, cheW, cheY,* and *cheZ;* and *cheV* and *cheR* respectively (44). Cluster II contains homologs of *cheA, cheB, cheW, cheY,* and *cheR,* however the functions of these genes have not been extensively studied (44). Genes in cluster III and IV are involved with cell aggregation associated with the "wrinkly spreader phenotype" and pilus-mediated twitching motility, respectively (44).

Out of 49 MCP, 35 contain the periplasmic LBD (4HB, HBM, PilJ, and CACHE), nine have the cytoplasmic LBD (PAS), one is comprised of the cytoplasmic NIT LBD, and four lack a LBD (43). One MCP (PSPTO_2480/PsPto-PscA) with the periplasmic CACHE domain has been

shown to be the sole chemoreceptor for L-aspartate, and is also able to sense both D-aspartate and L-glutamate (43). Both aspartate and glutamate are highly abundant in tomato apoplast, and the L-enantiomers of these amino acids can be used by *P. syringae* pv. tomato as carbon and nitrogen sources (45). Mutation of this *mcp* reduced *P. syringae* pv. tomato virulence in tomato and increased the level of cyclic-di-GMP, resulting in decreased swarming and increased biofilm formation (43). A recent study showed that mutations of two *mcp* (PSPTO_1008 and PSPTO_2526) in *P. syringae* pv. tomato DC3000 caused significant decrease in virulence compared to the wild-type in terms of symptom development and the number of colony forming units (46).

Ralstonia solanacearum

Ralstonia solanacearum is a Gram-negative, soil-borne pathogen responsible for bacterial wilt disease in more than 50 plant families (47). The typical route of infection for *R. solanacearum* is through plant roots via natural openings or wounds (47). The genome of *R. solanacearum* contains the chemotaxis genes *cheA, cheY, cheW, cheD, cheB,* and *cheZ,* and 21 MCP (47, 48). Eighteen MCP have the periplasmic LBD (4HB, PilJ and CACHE), two MCP contain the cytoplasmic LBD (PAS), and one has an unannotated LBD (48). The specific ligands that each *R. solanacearum* MCP sense are currently unknown, but both MCP with the PAS domain (Aer1 and Aer2) have been shown to be involved with energy taxis (aerotaxis) (49). Known chemo-attractants of *R. solanacearum* include amino acids (alanine, aspartate, asparagine, glutamate, glutamine, proline, and tyrosine), organic acids (citric acid, malic acid, and succinic acid) and secondary metabolites (4-hydroxycinnamic acid and 4-hydroxybenzoic acid) (47, 50). Furthermore, *R. solanacearum* was shown to be more attracted to the root exudates from its host plant tomato compared to the nonhost plant rice (47). The non-

chemotactic (*cheA* and *cheW*) mutants of *R. solanacearum* had similar reduction in symptom development in tomato plants compared to the wild-type strain, however, no difference in virulence was observed when the bacteria were inoculated directly into the plant stem (47). This suggests that chemotaxis is important for *R. solanacearum* in early stages of host infection for locating its host.

Xylella fastidiosa

Xylella fastidiosa is a Gram-negative, xylem-limited pathogen notorious for several diseases, most notably, Pierce's disease of grapes, citrus variegated chlorosis disease, and alfalfa dwarf disease (51). It can survive and multiply in the foregut of some leafhoppers such as the glassy-winged sharpshooter and the blue-green sharpshooter, which serve as vectors for this pathogen (51). When these insects probe plant tissues to locate and feed on the xylem, they acquire *X. fastidiosa* and transmit it to healthy plants (52). Interestingly, *X. fastidiosa* lacks flagella and its main mode of motility is twitching through the type IV pili where the bacteria translocate upstream against the flow of the transpiration stream (53). The genome of *X. fastidiosa* encodes one operon of six genes (*pilGIJL-chpBC*) that is homologous to the chemosensory genes (*cheYW-mcp-cheABW*) with only one MCP (48, 54). This operon is homologous to cluster IV chemosensory operon in *P. aeruginosa* that controls type IV twitching motility (54). The MCP in this cluster has an unannotated LBD and is presumed to be involved with twitching chemotaxis. The lack of MCP abundance in *X. fastidiosa* might be attributed to its relatively small niche in the plant xylem vessels compared to other plant pathogenic bacteria that live in fluctuating and diverse environment (54).

PRELIMINARY STUDIES

In 2014, multiple outbreaks of blackleg disease of potato occurred in the northeastern United States, resulting in significant crop losses (1, 55, 56). The main causal pathogen for these outbreaks was identified to be *D. dianthicola* (57). Since *Dickeya* can cause latent infections and infected seed potatoes still remain in the seed system, more outbreaks are expected to occur in the future; particularly considering that *D. dianthicola* has established itself in top seed potato-producing states such as Maine and Wisconsin (1, 56). One of the management methods for blackleg and soft rot disease is pathogen detection. During my first two years in Wisconsin, I was involved with testing hundreds of seed potatoes for the presence of *D. dianthicola* using polymerase chain reaction (PCR) assays. I discovered that one of the *D. dianthicola*-specific primers, DIA-A, failed to detect the presence of one of the isolates of *D. dianthicola* (designated TXG3 strain) (58). This finding has been published in a Disease Note of the Plant Disease journal (59). Furthermore, I am co-author for two publications where I contributed to showing that ME23 strain of *D. dianthicola* was a more typical isolate than TXG3 in the outbreak samples, and the hundreds of potato samples collected during the seed potato testing (57, 60).

Due to the prevalence of *D. dianthicola* in the United States, my research project initially focused on elucidating the function of chemotaxis in this species. *D. dianthicola* NCPPB 453 (type strain) genome encodes one copy of the chemotaxis genes *cheA, cheW, cheR, cheB, cheY, cheZ,* and *cheV,* and 42 MCP (61, 62). Using the Pfam database (63), the MCP architecture and the LBD were determined. Among 42 *D. dianthicola* MCP, 28 contain the 4HB domain, five consist of the HBM domain, three have the CACHE domain, one comprises of periplasmic FIST domain, three are made-up of the cytoplasmic PAS domain, one has the NIT domain, and one does not have any LBD (Figure 1.3). Most MCP with the periplasmic LBD contain the HAMP

domain, except for 20 MCP. However, the absence of the HAMP domain might be false negative due to the low sequence identity of the HAMP domains (10, 64).

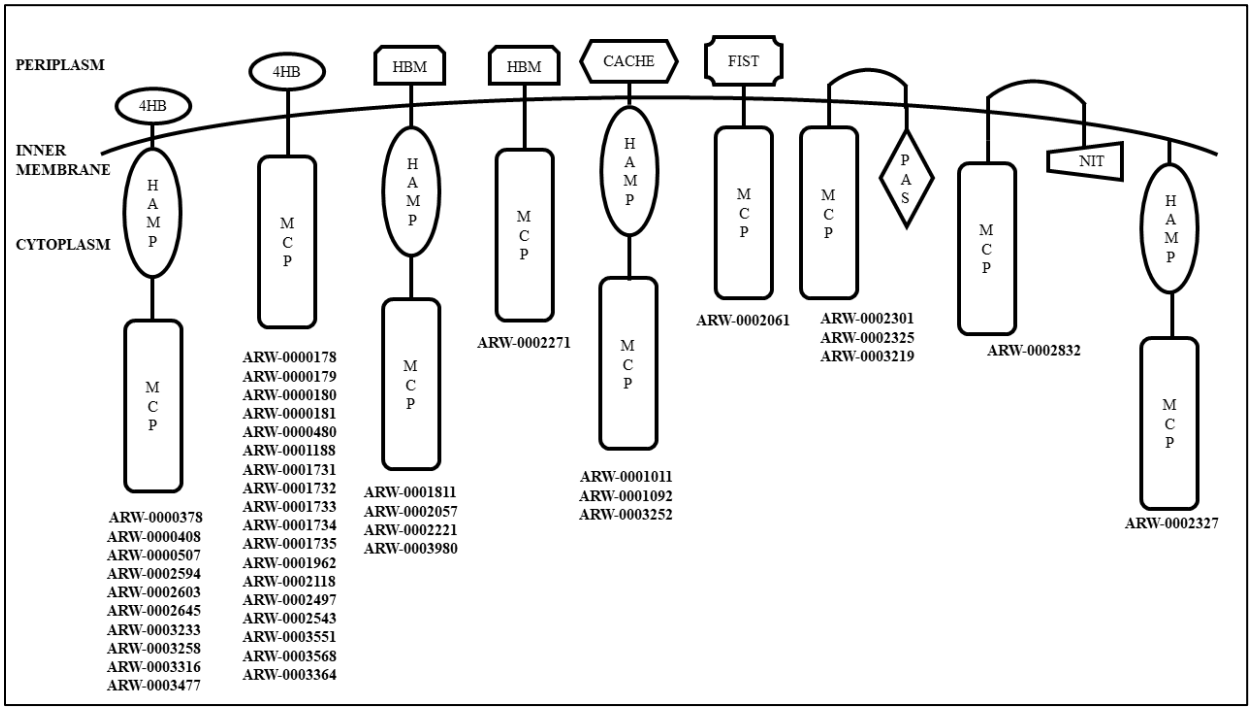

Figure 1.3: MCP architecture found in *D. dianthicola* NCPPB 453 genome. ASAP feature ID is provided for each corresponding MCP. 4HB: four-helix bundle domain; HAMP: histidine kinase, adenyl cyclase, methyl-accepting chemotaxis protein and phosphatase domain; MCP: MCP cytoplasmic signaling domain; HBM: helical bimodular domain; CACHE: calcium channels and chemotaxis receptors domain; FIST: F-box intracellular signal transduction domain; PAS: Per-Arnt-Sim proteins domain; NIT: nitrate and nitrite binding domain.

My next step was cloning MCP genes from *D. dianthicola* ME23 strain into *E. coli* UU1250 strain (has deletion of all five MCP genes of *E. coli*) to determine the function of each MCP. A total of 23 individual *D. dianthicola* MCP genes (Table 1.1) was successfully cloned into *E. coli* using the pGEM T-Easy vector plasmid (Promega, USA) based on the amplification of the corresponding MCP gene in the transformed cells. Minimal soft agar plate chemotaxis assay (0.25% [wt/vol] agar, 1 mM glycerol, 1 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 1 mM NaCl, 10 mM K_2HPO_4 at pH 7, 10 mM KH_2PO_4 at pH 7, 1 mg/l thiamine HCl, 0.1 mM L-histidine, 0.1 mM L-leucine, 0.1 mM L-methionine, and 0.1 mM L-threonine) was conducted to determine the

attractants for each individual MCP using previously described protocol (65). A volume of 10 uL of 100 mM concentration was used for the 20 chemicals tested (Table 1.2). *E. coli* RP437 (contains all 5 MCP of *E. coli*) was used as a positive control, and UU1250 was used as a negative control. Among the 23 MCP transformants, only one exhibited chemotactic ring (successive ring moving radially outwards from the inoculation point), which is ARW-0002301 that contains the PAS domain (aerotaxis transducer). This showed that the transformation worked. However, multiple attempts at determining what each of *D. dianthicola* MCP senses in *E. coli* transformants were unsuccessful as no chemotaxis phenotype was observed, except for the aerotaxis ortholog (ARW-0002301).

Table 1.1: List of 23 MCP genes of *D. dianthicola* cloned in *E. coli* UU1250 strain.

Ligand binding domain	**ASAP feature ID**
Four-helix bundle (4HB)	ARW-0000178 ARW-0001733 ARW-0000179 ARW-0001734 ARW-0000180 ARW-0001962 ARW-0000181 ARW-0002543 ARW-0000408 ARW-0002594 ARW-0001188 ARW-0002645 ARW-0001731 ARW-0003233 ARW-0001732
Helical bimodular (HBM)	ARW-0002057 ARW-0003980
Calcium channels and chemotaxis receptors (CACHE)	ARW-0001011 ARW-0003252 ARW-0001092
F-box intracellular signal transduction (FIST)	ARW-0002061
Per-Arnt-Sim (PAS)	ARW-0002301
-	ARW-0002327

Table 1.2: List of 20 chemicals tested in the plate chemotaxis assay.

Category	**Compounds tested**
Sugars	D-cellobiose, D-galactose, D-glucose, lactose, maltose, D-raffinose, L-rhamnose, D-ribose, sucrose
Amino acids	β-alanine, L-aspartic acid, L-glutamic acid, glycine, L-serine
Organic acids	Citric acid, D-galacturonic acid, L-malic acid
Plant hormones	3-indole acetic acid (auxin), benzyl adenine (cytokinin), methyl jasmonate (jasmonic acid)

There are several reasons that might explain why *D. dianthicola* MCP did not work as intended in *E. coli*:

1. *E. coli* is missing one chemotaxis protein, *cheV* in its genome. CheV functions as an adaptor protein (similar to CheW) that stabilizes the ternary complex between CheA and MCP (40). CheV has been shown to have an affinity bias towards specific MCP in which CheW is not capable of replacing CheV's function (40).
2. Chemotaxis by MCP can occur either through direct ligand binding, or via interaction with periplasmic binding protein (19-22). The proteins required for *D. dianthicola* MCP to sense specific ligands might not be present in *E. coli.*
3. MCP are arranged as trimers of dimers at the cell pole, and chemotactic response depends on the ratio of MCP (66). It could be that there was not enough ratio of the MCP to produce a chemotactic response in *E. coli*.

Afterward, overexpression of eight MCP genes (ARW-0000181, ARW-0001092, ARW-0001188, ARW-0001734, ARW-0002301, ARW-0002543, ARW-0002594, and ARW-0003980) was conducted in *D. dianthicola* ME23 using the pGEM T-Easy vector plasmid (Promega, USA). Following successful transformation and confirmation of corresponding MCP gene in ME23 transformants, minimal soft agar plate chemotaxis assay (0.25% [wt/vol] agar, 1 mM glycerol, 1 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 1 mM NaCl, 10 mM K_2HPO_4 at pH 7, and 10 mM KH_2PO_4 at pH 7) was carried out using the previously tested chemicals (Table 1.2). Wild-type ME23 strain was used as a control. Compared to the wild-type ME23 strain, all the transformants exhibited a much slower growth phenotype, leading to inconclusive results. Modifications to the composition of the soft agar were unsuccessful in getting the transformants to grow at relatively the same rate as the wild-type strain. It seemed that the overexpression of MCP using the high-

copy pGEM T-Easy vector plasmid caused growth defect in *D. dianthicola* ME23. This might be attributed to chemotaxis being energetically costly due to the use of ATP by CheA phosphorylation.

Minimal soft agar plate chemotaxis assays (0.25% [wt/vol] agar, 1 mM glycerol, 1 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 1 mM NaCl, 10 mM K_2HPO_4 at pH 7, and 10 mM KH_2PO_4 at pH 7) were also carried out for eight *Dickeya* strains and eight *Pectobacterium* strains (Table 1.3) to determine variations of chemotactic response. A total of 24 compounds (12 sugars, six amino acids, and six organic acids) at 100 nm concentration was tested in these assays. Furthermore, carbon utilization assay was done by replacing glycerol with 10 mM of each of the 24 compounds to test whether *Dickeya* and *Pectobacterium* can metabolize the compounds. The preliminary results of these assays were summarized in Table 1.4 (sugars), Table 1.5 (amino acids), and Table 1.6 (organic acids). No replications were carried out.

Table 1.3: List of 16 *Dickeya* and *Pectobacterium* strains used in the plate chemotaxis assay.

Species	**Strain**	**Description**
D. chrysanthemi (*Dc*)	0862	Isolated from corn
D. dadantii (*Dda*)	3937	Isolated from African violet
D. dianthicola (*Ddi*)	ME23	Isolated from potato in Wisconsin
	TXG3	Isolated from Texas potato sample, but most likely originated from Wisconsin
	IPO3646	Isolated in Netherlands
D. solani (*Ds*)	IPO2187	Isolated in Israel
	IPO3648	Isolated in Netherlands
D. zeae (*Dz*)	1591	Isolated from maize
P. atrosepticum (*Pa*)	CW1-1	Isolated from potato in Colorado
	CW1-4	Isolated from potato in Colorado
P. betavasculorum (*Pb*)	AC4150	Isolated from sugar beet
P. carotovorum subsp. *brasiliensis* (*Pcb*)	1692	Type strain; isolated from potato in Brazil
P. carotovorum subsp. *carotovorum* (*Pcc*)	ATCC 15713	Type strain
	WPP14	Isolated from Russet Norkotah stem in Wisconsin
P. parmentieri (*Pp*)	WPP163	Isolated from five infected potato tubers in Wisconsin
P. wasabiae (*Pw*)	ATCC 43316	Type strain; isolated from *Eutrema wasabi* in Nagano Prefecture, Japan

Based on Table 1.4, all strains had chemotactic response towards D-ribose and use it as a sole carbon source. There were variations in chemotactic response between strains towards D-glucose, sucrose, D-fructose, and D-galactose, but all strains can metabolize these sugars. Glucose, fructose, and sucrose were found to be rapidly used by SRP during infection (67). None of the strains had attractant response towards D-cellobiose, D-raffinose, and maltose despite being able to use them as carbon sources. L-rhamnose, lactose, myo-inositol, and D-trehalose were not chemoeffectors for any of the strains, and there were variations in utilizations of these sugars as carbon sources.

In Table 1.5, all strains showed chemotaxis towards L-aspartic acid, which could not be utilized as a sole carbon source. Chemoattraction towards L-aspartic acid is not surprising considering that it is the most abundant amino acid in the plants (68). Besides *D. dadantii* 3937, the remaining 15 strains showed chemotaxis towards L-asparagine that was not a usable carbon source. Asparagine is an efficient molecule for nitrogen storage and transport in plants (69), and nitrogen abundance is associated with increased disease severity in plants (70). This implies that asparagine could be a signal for plant pathogens to find susceptible host plants. *R. solanacearum*, a soilborne plant pathogen, was shown to be chemotactically attracted to L-asparagine (47). Among the nine strains that exhibited chemotaxis towards serine, five strains (*D. chrysanthemi* 0862*, D. zeae* 1591, *P. betavasculorum* AC4150, *P. carotovorum* subsp. *brasiliensis* 1692, and *P. carotovorum* subsp. *carotovorum* ATCC 15713) were able to metabolize serine. Few strains were attracted to L-glutamic acid, glycine, and β-alanine; however, most strains could not metabolize these amino acids.

For organic acids (Table 1.6), all *Dickeya* and *Pectobacterium* strains had chemotactic response towards L-malic acid, but only *Dickeya* strains could metabolize it. L-malic acid is one

of the compounds secreted by plant roots to recruit beneficial rhizobacteria (71). This suggests that *Dickeya* and *Pectobacterium* use L-malic acid as a signal to find their host plants in the soil. Despite the variations in chemotactic response towards D-galacturonic acid and pyruvic acid, all strains could use them as carbon sources. Variations in chemotactic response towards citric acid and sodium citrate were observed. Interestingly, all *Pectobacterium* strains did not exhibit chemotactic response towards citric acid. Moreover, none of the strains could metabolize citric acid. All strains did not show chemotactic response towards maleic acid nor could use it as a carbon source.

From these preliminary chemotaxis results, I proceeded to determine the abundance of MCP in soft rot Pectobacteriaceae compared to other relevant bacteria in the order Enterobacteriales, which is summarized in Chapter 2.

Besides identifying specific ligands that *D. dianthicola* MCP sense, my project aims also include determining how the MCP genes are regulated. The FliA sigma factor has been found to regulate the initiation of MCP genes transcription in *E. coli* (72), and its binding motif has been identified as TAAAGTTT-N11-GCCGATAA (73). The 500-bp sequences upstream of 42 MCP genes of *D. dianthicola* NCPPB 453 were obtained from the ASAP database (61, 62). Manual identification of FliA motif was carried out for each MCP. Among 42 *D. dianthicola* MCP genes, 22 have putative FliA motif upstream (Table 1.7). Mutation of *fliA* in *D. dianthicola* ME23 strain was attempted multiple times using two methods: 1) phage lambda red recombinase system (74) and 2) splicing by overlapping extension PCR, however, to no avail. Therefore, I decided to use the available *fliA* mutant of *D. dadantii* 3937 made by Dr. Courtney Jahn (8) for the RNA-sequencing experiment in Chapter 3.

Table 1.4: Summary of chemotactic and metabolic response of 16 *Dickeya* and *Pectobacterium* strains to 12 sugars.

Sugars		***Dc* 0862**	***Dda* 3937**	***Ddi* ME23**	***Ddi* TXG3**	***Ddi* 3646**	***Ds* 2187**	***Ds* 3648**	***Dz* 1591**	***Pa* 1-1**	***Pa* 1-4**	***Pb* 4150**	***Pcb* 1692**	***Pcc* 15713**	***Pcc* 14**	***Pp* 163**	***Pw* 43316**
D-ribose		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
D-glucose		+	+	+	+	+	+	+	+	+	+	-	+	+	-	+	+
Sucrose		+	+	+	-	-	-	+	+	+	+	+	+	+	+	+	+
D-fructose		-	-	+	-	+	-	+	-	+	+	+	+	-	-	+	+
D-galactose		+	-	+	+	+	+	+	+	-	-	-	-	-	+	+	+
D-cellobiose		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D-raffinose		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Maltose		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L-rhamnose		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lactose		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Myo-inositol		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D-trehalose		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Abbreviations for each strain were based on Table 3. Only the strain number was provided due to limited space.
'+' indicates chemotaxis response; '-' indicates no chemotaxis response.
Gray color indicates that the compound can be metabolized as a sole carbon source; white color indicates otherwise.

Table 1.5: Summary of chemotactic and metabolic response of 16 *Dickeya* and *Pectobacterium* strains to six amino acids.

Sugars	***Dc* 0862**	***Dda* 3937**	***Ddi* ME23**	***Ddi* TXG3**	***Ddi* 3646**	***Ds* 2187**	***Ds* 3648**	***Dz* 1591**	***Pa* 1-1**	***Pa* 1-4**	***Pb* 4150**	***Pcb* 1692**	***Pcc* 15713**	***Pcc* 14**	***Pp* 163**	***Pw* 43316**
L-aspartic acid	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
L-asparagine	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
L-serine	+	+	-	-	+	-	+	+	-	-	+	+	+	+	-	-
L-glutamic acid	+	-	-	-	-	+	-	+	-	-	-	+	+	+	-	-
Glycine	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-
B-alanine	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-

Abbreviations for each strain are based on Table 3. Only the strain number was provided due to limited space.
'+' indicates chemotaxis response; '-' indicates no chemotaxis response.
Gray color indicates that the compound can be metabolized as a sole carbon source; white color indicates otherwise.

Table 1.6: Summary of chemotactic and metabolic response of 16 *Dickeya* and *Pectobacterium* strains to six organic acids.

Sugars	***Dc* 0862**	***Dda* 3937**	***Ddi* ME23**	***Ddi* TXG3**	***Ddi* 3646**	***Ds* 2187**	***Ds* 3648**	***Dz* 1591**	***Pa* 1-1**	***Pa* 1-4**	***Pb* 4150**	***Pcb* 1692**	***Pcc* 15713**	***Pcc* 14**	***Pp* 163**	***Pw* 43316**
L-malic acid	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
D-galacturonic acid	+	-	-	+	-	+	+	+	+	+	-	-	+	+	+	+
Citric acid	+	-	+	+	+	-	+	+	-	-	-	-	-	-	-	-
Sodium citrate	-	-	-	-	-	-	-	+	-	-	+	+	-	-	-	-
Pyruvic acid sodium salt	-	-	-	-	-	-	+	+	-	-	+	+	+	+	+	+
Maleic acid	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Abbreviations for each strain are based on Table 3. Only the strain number was provided due to limited space.
'+' indicates chemotaxis response; '-' indicates no chemotaxis response.
Gray color indicates that the compound can be metabolized as a sole carbon source; white color indicates otherwise.

Table 1.7: Putative FliA binding sites upstream of *D. dianthicola* NCPPB 453 MCP genes compared to *E. coli* K-12 sub-strain MG1655 MCP genes. The consensus FliA motif consists of TAAAGTTT-N11-GCCGATAA (73).

ASAP Feature ID	**Putative FliA Motif**
***E. coli* K-12 substr. MG1655**	
ABE-0004744 (*trg*)	**TAAGT**AA**T**taccgtcaagt**GCCGAT**G**A**C
ABE-0006290 (*tar-tap*)	**TAAAGTTT**ccccctcctt**GCCGATAA**C
ABE-0010087 (*aer*)	**TAAAG**A**T**Aaccgcagcggg**GCCGA**C**A**TA
ABE-0014282 (*tsr*)	**TAAAGTTT**ttcctttccag**GCCGA**A**AA**T
***D. dianthicola* NCPPB 453**	
ARW-0003980	**TAAAGTTT**cactatggcgt**GCCGATA**GA
ARW-0003551	**TAAA**TAAAaagacggaatt**GCCGATA**TA
ARW-0003233	**TAAAG**AACaacggcagtatT**CCGATAA**G
ARW-0003252	**TAAA**A**TTT**cgcacttttt**GCCG**T**TAA**C
ARW-0002327	**TAAAG**GCAgccggtcgttg**G**T**CGAT**T**A**A
ARW-0000507	**TAA**TC**TT**Cgatgttgacac**GCCGATAA**A
ARW-0002118	**T**C**AAGA**T**G**ccagttttact**G**T**CGATA**TT
ARW-0002061	**TA**CC**G**GCGgcgataactac**GCCGATA**CG
ARW-0001092	**TAAA**AGG**T**gatcaaacaat**G**AA**GATA**TC
ARW-0001188	**TAAA**CA**TT**cctatagtttt**G**A**CGAT**G**A**C
ARW-0002594	**TAAAG**G**T**GttaccaatcttCG**CG**G**T**GTC
ARW-0000180	**TAAA**T**T**AAtccccggaatcA**C**A**GATAA**T
ARW-0002645	**TAAA**CGGCtaccgacgatgA**C**A**GA**A**A**TA
ARW-0000378	**TAAA**T**T**GAattattcagtaAT**CGA**CT**A**T
ARW-0002543	**TAAA**T**T**AAatatatcgctgTAA**GATAA**T
ARW-0000178	**T**CCT**G**CGGctcgaattattTAG**GATA**TA
ARW-0002301	**T**T**A**GC**T**CAtcaatatttccTG**CGATAA**A
ARW-0003364	**T**GTTCAAAttagctgcgtc**G**AT**GATA**TT
ARW-0001011	**TAAA**TC**T**Gtaaatctgata**GC**A**G**GC**A**TT
ARW-0002497	**T**GC**A**A**T**GGaatgattttctTAA**GATA**TG
ARW-0001811	A**AA**TC**T**C**T**atgctgcaagtA**CCGATAA**C
ARW-0001733	C**A**TT**G**AACgtgttgccggtCAT**GATA**GA

REFERENCES

1. Charkowski AO. The changing face of bacterial soft-rot diseases. Annual Review of Phytopathology. 2018;56:269-88.

2. Adeolu M, Alnajar S, Naushad S, Gupta RS. Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae, Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. International Journal of Systematic and Evolutionary Microbiology. 2016;66(12):5575-99.

3. Reverchon S, Nasser W. *Dickeya* ecology, environment sensing and regulation of virulence programme. Environmental Microbiology Reports. 2013;5(5):622-36.

4. Ma B, Hibbing ME, Kim H-S, Reedy RM, Yedidia I, Breuer J, et al. Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. Phytopathology. 2007;97(9):1150-63.

5. Charkowski A, Blanco C, Condemine G, Expert D, Franza T, Hayes C, et al. The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. Annual Review of Phytopathology. 2012;50:425-49.

6. Kim H-S, Thammarat P, Lommel SA, Hogan CS, Charkowski AO. *Pectobacterium carotovorum* elicits plant cell death with DspE/F but the *P. carotovorum* DspE does not suppress callose or induce expression of plant genes early in plant–microbe interactions. Molecular Plant-Microbe Interactions. 2011;24(7):773-86.

7. Hossain MM, Shibata S, Aizawa S-I, Tsuyumu S. Motility is an important determinant for pathogenesis of *Erwinia carotovora* subsp. *carotovora*. Physiological and Molecular Plant Pathology. 2005;66(4):134-43.

8. Jahn CE, Willis DK, Charkowski AO. The flagellar sigma factor FliA is required for *Dickeya dadantii* virulence. Molecular Plant-Microbe Interactions. 2008;21(11):1431-42.

9. Marquez-Villavicencio MdP, Groves RL, Charkowski AO. Soft rot disease severity is affected by potato physiology and *Pectobacterium* taxa. Plant Disease. 2011;95(3):232-41.

10. Ud-Din AIMS, Roujeinikova A. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. Cellular and Molecular Life Sciences. 2017;74(18):3293-303.

11. Matilla MA, Krell T. The effect of bacterial chemotaxis on host infection and pathogenicity. FEMS Microbiology Reviews. 2017;42(1):fux052.

12. Bansal T, Englert D, Lee J, Hegde M, Wood TK, Jayaraman A. Differential effects of epinephrine, norepinephrine, and indole on *Escherichia coli* O157: H7 chemotaxis, colonization, and gene expression. Infection and Immunity. 2007;75(9):4597-607.

13. Antunez-Lamas M, Cabrera E, Lopez-Solanilla E, Solano R, González-Melendi P, Chico JM, et al. Bacterial chemoattraction towards jasmonate plays a role in the entry of *Dickeya dadantii* through wounded tissues. Molecular Microbiology. 2009;74(3):662-71.

14. Wuichet K, Zhulin IB. Origins and diversification of a complex signal transduction system in prokaryotes. Science Signaling. 2010;3(128):ra50-ra.

15. Briegel A, Ortega DR, Tocheva EI, Wuichet K, Li Z, Chen S, et al. Universal architecture of bacterial chemoreceptor arrays. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(40):17181-6.

16. Liu J, Hu B, Morado DR, Jani S, Manson MD, Margolin W. Molecular architecture of chemoreceptor arrays revealed by cryoelectron tomography of *Escherichia coli* minicells.

Proceedings of the National Academy of Sciences of the United States of America. 2012;109(23):E1481-E8.

17. Milburn MV, Prive GG, Milligan DL, Scott WG, Yeh J, Jancarik J, et al. Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. Science. 1991;254(5036):1342-7.

18. Henry JT, Crosson S. Ligand-binding PAS domains in a genomic, cellular, and structural context. Annual Review of Microbiology. 2011;65:261-86.

19. Hugouvieux-Cotte-Pattat N, Blot N, Reverchon S. Identification of TogMNAB, an ABC transporter which mediates the uptake of pectic oligomers in *Erwinia chrysanthemi* 3937. Molecular Microbiology. 2001;41(5):1113-23.

20. Konishi H, Hio M, Kobayashi M, Takase R, Hashimoto W. Bacterial chemotaxis towards polysaccharide pectin by pectin-binding protein. Scientific Reports. 2020;10.

21. Hegde M, Englert DL, Schrock S, Cohn WB, Vogt C, Wood TK, et al. Chemotaxis to the quorum-sensing signal AI-2 requires the Tsr chemoreceptor and the periplasmic LsrB AI-2-binding protein. Journal of Bacteriology. 2011;193(3):768-73.

22. Gardina P, Conway C, Kossman M, Manson M. Aspartate and maltose-binding protein interact with adjacent sites in the Tar chemotactic signal transducer of *Escherichia coli*. Journal of Bacteriology. 1992;174(5):1528-36.

23. Ortega Á, Zhulin IB, Krell T. Sensory repertoire of bacterial chemoreceptors. Microbiology and Molecular Biology Reviews. 2017;81(4).

24. Lacal J, García-Fontana C, Muñoz-Martínez F, Ramos JL, Krell T. Sensing of environmental signals: classification of chemoreceptors according to the size of their ligand binding regions. Environmental Microbiology. 2010;12(11):2873-84.

25. Mougel C, Zhulin IB. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends in Biochemical Sciences. 2001;26(10):582-4.

26. Anantharaman V, Aravind L. Cache–a signaling domain common to animal Ca2+-channel subunits and a class of prokaryotic chemotaxis receptors. Trends in Biochemical Sciences. 2000;25(11):535-7.

27. Ulrich LE, Zhulin IB. Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. Bioinformatics. 2005;21(Suppl_3):iii45-iii8.

28. Ortega Á, Krell T. The HBM domain: introducing bimodularity to bacterial sensing. Protein Science. 2014;23(3):332-6.

29. Shu CJ, Ulrich LE, Zhulin IB. The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors. Trends in Biochemical Sciences. 2003;28(3):121-4.

30. Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiology and Molecular Biology Reviews. 1999;63(2):479-506.

31. Bibikov SI, Miller AC, Gosink KK, Parkinson JS. Methylation-independent aerotaxis mediated by the *Escherichia coli* Aer protein. Journal of bacteriology. 2004;186(12):3730-7.

32. Stewart V. Nitrate regulation of anaerobic respiratory gene expression in *Escherichia coli*. Molecular Microbiology. 1993;9(3):425-34.

33. Borziak K, Zhulin IB. FIST: a sensory domain for diverse signal transduction pathways in prokaryotes and ubiquitin signaling in eukaryotes. Bioinformatics. 2007;23(19):2518-21.

34. Wuichet K, Alexander RP, Zhulin IB. Comparative genomic and protein sequence analyses of a complex system controlling bacterial chemotaxis. Methods in Enzymology. 422: Elsevier; 2007. p. 3-31.

35. Liu X, Parales RE. Chemotaxis of *Escherichia coli* to pyrimidines: a new role for the signal transducer tap. Journal of Bacteriology. 2008;190(3):972-9.

36. Manson MD, Blank V, Brade G, Higgins CF. Peptide chemotaxis in *E. coli* involves the Tap signal transducer and the dipeptide permease. Nature. 1986;321(6067):253-6.

37. Lin L-N, Li J, Brandts JF, Weis RM. The serine receptor of bacterial chemotaxis exhibits half-site saturation for serine binding. Biochemistry. 1994;33(21):6564-70.

38. Kondoh H, Ball CB, Adler J. Identification of a methyl-accepting chemotaxis protein for the ribose and galactose chemoreceptors of *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America. 1979;76(1):260-4.

39. Parkinson JS, Hazelbauer GL, Falke JJ. Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. Trends in Microbiology. 2015;23(5):257-66.

40. Ortega DR, Zhulin IB. Evolutionary genomics suggests that CheV is an additional adaptor for accommodating specific chemoreceptors within the chemotaxis signaling complex. PLOS Computational Biology. 2016;12(2):e1004723.

41. Sherris D, Parkinson JS. Posttranslational processing of methyl-accepting chemotaxis proteins in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America. 1981;78(10):6051-5.

42. Clarke CR, Hayes BW, Runde BJ, Markel E, Swingle BM, Vinatzer BA. Comparative genomics of *Pseudomonas syringae* pathovar tomato reveals novel chemotaxis pathways associated with motility and plant pathogenicity. PeerJ. 2016;4:e2570.

43. Cerna-Vargas JP, Santamaría-Hernando S, Matilla MA, Rodríguez-Herva JJ, Daddaoua A, Rodríguez-Palenzuela P, et al. Chemoperception of specific amino acids controls phytopathogenicity in *Pseudomonas syringae* pv. tomato. mBio. 2019;10(5).

44. Parales RE, Ferrández A, Harwood CS. Chemotaxis in pseudomonads. Pseudomonas: Springer; 2004. p. 793-815.

45. Rico A, Preston GM. *Pseudomonas syringae* pv. tomato DC3000 uses constitutive and apoplast-induced nutrient assimilation pathways to catabolize nutrients that are abundant in the tomato apoplast. Molecular Plant-Microbe Interactions. 2008;21(2):269-82.

46. Santamaría-Hernando S, Cerna-Vargas JP, Martínez-García PM, de Francisco-de Polanco S, Nebreda S, Rodríguez-Palenzuela P, et al. Blue-light perception by epiphytic *Pseudomonas syringae* drives chemoreceptor expression, enabling efficient plant infection. Molecular Plant Pathology. 2020;21(12):1606-19.

47. Yao J, Allen C. Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*. Journal of Bacteriology. 2006;188(10):3697-708.

48. Gumerov VM, Ortega DR, Adebali O, Ulrich LE, Zhulin IB. MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. Nucleic Acids Research. 2020;48(D1):D459-D64.

49. Yao J, Allen C. The plant pathogen *Ralstonia solanacearum* needs aerotaxis for normal biofilm formation and interactions with its tomato host. Journal of Bacteriology. 2007;189(17):6415-24.

50. Hasegawa T, Kato Y, Okabe A, Itoi C, Ooshiro A, Kawaide H, et al. Effect of secondary metabolites of tomato (*Solanum lycopersicum*) on chemotaxis of *Ralstonia solanacearum*, pathogen of bacterial wilt disease. Journal of Agricultural and Food Chemistry. 2019;67(7):1807-13.

51. Hopkins D, Purcell A. *Xylella fastidiosa*: cause of Pierce's disease of grapevine and other emergent diseases. Plant Disease. 2002;86(10):1056-66.

52. Purcell AH, Hopkins DL. Fastidious xylem-limited bacterial plant pathogens. Annual Review of Phytopathology. 1996;34(1):131-51.

53. Meng Y, Li Y, Galvani CD, Hao G, Turner JN, Burr TJ, et al. Upstream migration of *Xylella fastidiosa* via pilus-driven twitching motility. Journal of Bacteriology. 2005;187(16):5560-7.

54. Cursino L, Galvani CD, Athinuwat D, Zaini PA, Li Y, De La Fuente L, et al. Identification of an operon, Pil-Chp, that controls twitching motility and virulence in *Xylella fastidiosa*. Molecular Plant-Microbe Interactions. 2011;24(10):1198-206.

55. Ma X, Schloop A, Swingle B, Perry KL. *Pectobacterium* and *Dickeya* responsible for potato blackleg disease in New York State in 2016. Plant Disease. 2018;102(9):1834-40.

56. Jiang H, Hao J, Johnson S, Brueggeman R, Secor G. First report of *Dickeya dianthicola* causing blackleg and bacterial soft rot on potato in Maine. Plant Disease. 2016;100(11):2320-.

57. Ma X, Perna NT, Glasner JD, Hao J, Johnson S, Nasaruddin AS, et al. Complete Genome Sequence of *Dickeya dianthicola* ME23, a Pathogen Causing Blackleg and Soft Rot Diseases of Potato. Microbiology Resource Announcements. 2019;8(7).

58. Pritchard L, Humphris S, Saddler G, Parkinson N, Bertrand V, Elphinstone J, et al. Detection of phytopathogens of the genus *Dickeya* using a PCR primer prediction pipeline for draft bacterial genome sequences. Plant Pathology. 2013;62(3):587-96.

59. Nasaruddin A, Charkowski A, Babler B, Perna N, Glasner J. First report of *Dickeya dianthicola* causing blackleg on potato in Texas. Plant Disease. 2019;103(8):2121-.

60. Karim S, McNally RR, Nasaruddin AS, DeReeper A, Mauleon RP, Charkowski AO, et al. Development of the automated primer design workflow uniqprimer and diagnostic primers for the Broad-Host-Range plant pathogen *Dickeya dianthicola*. Plant Disease. 2019;103(11):2893-902.

61. Glasner JD, Liss P, Plunkett III G, Darling A, Prasad T, Rusch M, et al. ASAP, a systematic annotation package for community analysis of genomes. Nucleic Acids Research. 2003;31(1):147-51.

62. Glasner JD, Rusch M, Liss P, Plunkett III G, Cabot EL, Darling A, et al. ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Research. 2006;34(suppl_1):D41-D5.

63. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Research. 2019;47(D1):D427-D32.

64. Appleman JA, Stewart V. Mutational analysis of a conserved signal-transducing element: the HAMP linker of the *Escherichia coli* nitrate sensor NarX. Journal of Bacteriology. 2003;185(1):89-97.

65. Parkinson JS. Phenol sensing by *Escherichia coli* chemoreceptors: a nonclassical mechanism. Journal of Bacteriology. 2011;193(23):6597-604.

66. Kalinin Y, Neumann S, Sourjik V, Wu M. Responses of *Escherichia coli* bacteria to two opposing chemoattractant gradients depend on the chemoreceptor ratio. Journal of Bacteriology. 2010;192(7):1796-800.

67. Effantin G, Rivasseau C, Gromova M, Bligny R, Hugouvieux-Cotte-Pattat N. Massive production of butanediol during plant infection by phytopathogenic bacteria of the genera *Dickeya* and *Pectobacterium*. Molecular Microbiology. 2011;82(4):988-97.

68. Kumar V, Sharma A, Kaur R, Thukral AK, Bhardwaj R, Ahmad P. Differential distribution of amino acids in plants. Amino Acids. 2017;49(5):821-69.

69. Lea PJ, Sodek L, Parry MA, Shewry PR, Halford NG. Asparagine in plants. Annals of Applied Biology. 2007;150(1):1-26.

70. Huber D, Watson R. Nitrogen form and plant disease. Annual Review of Phytopathology. 1974;12(1):139-65.

71. Rudrappa T, Czymmek KJ, Paré PW, Bais HP. Root-secreted malic acid recruits beneficial soil bacteria. Plant Physiology. 2008;148(3):1547-56.

72. Fitzgerald DM, Bonocora RP, Wade JT. Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. PLOS Genetics. 2014;10(10):e1004649.

73. Ide N, Ikebe T, Kutsukake K. Reevaluation of the promoter structure of the class 3 flagellar operons of *Escherichia coli* and *Salmonella*. Genes and Genetic Systems. 1999;74(3):113-6.

74. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(12):6640-5.

Chapter 2

Comparative genomic analyses of the order Enterobacteriales support the importance of ATP-binding cassette transporters and methyl-accepting chemotaxis proteins in the pathogenicity of soft rot Pectobacteriaceae

SUMMARY

The ATP-binding cassette (ABC) transporters are membrane proteins crucial for importing and exporting various compounds across cell membranes. Some ABC transporters interact with methyl-accepting chemotaxis proteins (MCP), chemoreceptors that alter the direction of bacterial flagellar rotation by sensing environmental and intracellular signals. The wealth of fully sequenced prokaryotic genomes allows us to determine the abundance of ABC transporters and MCP in the order Enterobacteriales, which consists of notable pathogenic and non-pathogenic bacteria such as *Escherichia coli, Yersinia, Pectobacterium, Erwinia*, and *Sodalis*. Compared to non-pathogenic and animal pathogenic bacteria in the order, soft rot Pectobacteriaceae, namely *Dickeya* and *Pectobacterium*, encode more ABC transporters and MCP in their genomes. For ABC transporters, soft rot Pectobacteriaceae genomes are enriched with sugar, peptide, amino acid, iron-siderophore, and family 5 transporters, in addition to family 6 and 9 exporters, compared to other bacteria in the order. For MCP, soft rot Pectobacteriaceae not only contain more than 30 MCP genes per strain, but also have more diverse ligand binding domains than other species in the Enterobacteriales. The findings from this comparative genomic study suggest that both ABC transporters and MCP are important for soft rot Pectobacteriaceae pathogenicity, that enrichment of these two systems may be coupled and may be an important

step in the evolution of soft rot pathogenesis, and that these systems could be the targets for plant disease management.

INTRODUCTION

ABC transporters use ATP hydrolysis to actively import or export diverse molecules across cell membranes, including sugars, peptides, amino acids, ions, antibiotics, and toxins (1). In addition to their role as gatekeepers, a subset of ABC transporters interacts with MCP (2-4), which are chemoreceptors that participate in bacterial chemotaxis, a process that guides cell movement in response to environmental and intracellular signals via control of flagellar rotation (5, 6). Therefore, bacteria use MCP alone or together with ABC transporters to sense and respond to fluctuating environments, suggesting a cooperative role of MCP and ABC transporters in host adaptation and pathogenicity (7).

Both ABC transporters and MCP have unique domain features. ABC transporters are often comprised of a nucleotide-binding domain(s) (NBD) and a transmembrane domain(s) (TMD), and the ABC importers require an additional substrate-binding protein (SBP). Of the different domain components in an ABC transporter operon, the nucleotide sequence of the NBD domain, which functions as an ATPase to supply energy to either import or export molecules against a concentration gradient, is the most conserved. Thus, ATPase homology is commonly used to categorize ABC transporter families (8-10). However, some aspects of ABC transporters are modular and are not aligned with ATPase homology, such as the type of SBP encoded by importers (11).

Like ABC transporters, MCP are modular, complicating categorization and phylogeny of this protein family. MCP consist of up to four main domains: a ligand binding domain (periplasmic or cytoplasmic), transmembrane helices, a histidine kinase-adenyl cyclase-MCP-

phosphatase (HAMP) linker domain, and cytoplasmic signaling domain (5). MCP with periplasmic ligand binding domains may detect extracellular signals, whereas MCP with cytoplasmic ligand binding domains detect intracellular signals, at least some of which are correlated with the energy state of the cell (5). Ligand binding alters MCP conformation, which relays the signal to the cytoplasmic signaling domain through transmembrane helices and the HAMP linker domain (5). Besides MCP, there are several core proteins involved in the chemotaxis signal transduction system in prokaryotes, namely CheA, CheW, CheY, CheR, CheB, and CheZ (12, 13). Among the four domains that compose MCP, the ligand binding domain is the least conserved domain (12). In contrast, the cytoplasmic signaling domain has the highest degree of sequence conservation due to the requirement that it maintains interaction with CheA and CheW (12).

The number of ABC transporters and MCP encoded in bacterial genomes may be attributed to their lifestyles and environments. We hypothesized that bacteria living in diverse, fluctuating environments encode more ABC transporters and MCP, which allow them to sense, import or export more compounds. Ultimately, the high number of ABC transporters and MCP confer competitive fitness advantage in nutrient allocation and utilization, and the ability to adapt to multiple environments.

Most work in bacteria with ABC transporters and MCP has been completed with the model species *Escherichia coli*, which is in the Enterobacteriales. The genomes of nearly all type strains in Enterobacteriales are sequenced, which make this order a good model for studying the abundance of ABC transporters and MCP and determining if there are correlations between transporter or MCP content with the natural history of each species. This order of Gram-negative bacteria contains seven families, including the Budviciaceae, Enterobacteriaceae, Erwiniaceae,

Hafniaceae, Morganellaceae, Pectobacteriaceae, and Yersiniaceae (14), and includes numerous well-known non-pathogenic and pathogenic bacterial species that colonize soil, water, plants, and animals.

In this study, we investigated the number of ABC transporters and MCP encoded in Enterobacteriales, using type strains when possible, determined homologous relationships of ABC members among different bacterial species, examined the phylogeny of ABC transporters and MCP, and identified ABC transporters likely to interact with MCP. We also compared the domain features in ABC transporters and MCP to further our understanding of their roles in phytopathogenicity.

MATERIALS AND METHODS

Genome sources and strains used in this study

We primarily analyzed genomes of type strains for each family within the Enterobacteriales (14). We also included type strains that are not in the type genus (i.e. genus that represents the typical characteristics of the family it belongs to) but that are well-annotated, including strains within *Dickeya, Brenneria*, *Lonsdalea*, *Sodalis,* and *Salmonella*, to better represent the diversity within this order. Thus, a total of 72 type strains consisting of animal pathogens, plant pathogens, and non-pathogenic bacteria that inhabit different environments, were chosen for the comparative genome analyses (Table 2.1). All sequences were retrieved from the ASAP database (15, 16), but not all genomes were used in every analysis conducted.

Ortholog identification of nucleotide-binding domains of ABC transporters

To identify orthologous relationship of the NBD of ABC transporters among the 70 organisms (excluding *Salmonella* Typhimurium LT2 and *Sodalis praecaptivus* ATCC BAA-2554), we first extracted 108 NBD sequences, including 106 ABC transporters in

Pectobacterium atrosepticum SCRI 1043, which has the best annotated genome in plant pathogens in Enterobacteriales, and two other ABC transporters in *E. coli* K-12 MG1655 to represent ten ABC transporter families (Table 2.2) (8). Each NBD sequence was used as the query to perform the BLASTP searches against all other genomes for amino acid sequence similarity comparisons, following three criteria, including 1) a score > 300 bits, 2) amino acid sequences with > 35% identity and > 65% of the length, and 3) an *e*-value < 1×10^{-80}. When the similarity criteria were satisfied, the presence of TMD and/or SBD at either upstream or downstream of the NBD were manually verified to identify protein orthologs. The presence of each ABC transporter was then manually recorded.

Phylogenetic analysis and sequence conservation of the NBDs

NBD sequence characteristics of ABC transporter are the main criteria for ABC transporter classification. To construct a phylogenetic tree showing relationships among NBDs of every ABC transporter member, the 108 NBD sequences were aligned using ClustalW algorithm (17) with gap open penalty of 10 and gap extension penalty of 0.2. The resulted sequence matrix was subjected to a phylogeny analysis using a maximum likelihood method based on the JTT protein substitution model with 500 bootstrap replicates to generate a phylogenetic tree in MEGA X (18). The conservation of amino acid sequences of NBD that belong to the same family was investigated in a way similar to that the NBD used for the phylogenetic analysis, and the motifs were examined using the search engines Motif Scan (19) and InterPro (20). The sequence conservation and phylogenetic analysis were performed with the aid of R packages including msa, ape, phangorn, and ggtree.

ABC transporter diversity and gene structures

Of the 70 type strains, we further selected 12 strains to show phylogenetic relations of these bacteria in Enterobacteriales by comparing the housekeeping gene *dnaX* sequences. We included *P. atrosepticum* SCRI 1043, *P. carotovorum* subsp. *carotovorum* ATCC 15713, *Brenneria salicis* ATCC 15712, *Dickeya chrysanthemi* ATCC 11633, *Lonsdalea quercina* ATCC 29281, *Yersinia pestis* ATCC 19428, *Hafnia alvei* ATCC 13337, *Erwinia amylovora* ATCC 15580, *Enterobacter cloacae* subsp. *cloacae* ATCC 13047, *Escherichia coli* ATCC 11775, *Morganella morganii* subsp. *morganii* ATCC 25830, and *Budvicia aquatica* ATCC 35567. The number of ABC transporters in each family was summed to aid in comparing ABC transporter diversity among the selected bacteria. Like eukaryote ABC transporters, prokaryotic transporters display different operon structures in a bacterial genome. Thus, the operon structures of ABC transporter paralogs were compared to understand the difference of gene structures in Enterobacteriales.

Phylogenetic analysis of MCP signaling domain and determination of the chemotaxis system

To determine the evolutionary relationships among MCP, we first retrieved a total of 217 MCP protein sequences (Table 2.3) from 14 strains (Table 2.4) to represent the seven families in the order Enterobacteriales. All five genera in Pectobacteriaceae were included since they contain important plant pathogens and non-pathogenic bacteria. We chose to use MCP protein sequences from *Escherichia coli* K-12 MG1655 rather than the *E. coli* type strain because it contains five MCP (Tsr, Tar, Tap, Trg, and Aer) and the type strain only has three MCP. The signaling domain protein sequences were identified by searching the 217 MCP protein sequences against the SMART database in 'genomic mode' for methyl-accepting (MA) domain model (21). The alignment of MCP signaling domain protein sequences was conducted using the same

criteria as ABC transporter, and a maximum likelihood phylogenetic tree was built in R as described above. Furthermore, we used the MiST database to determine the number of chemotaxis systems present in the 14 bacterial strains (22).

Identification of MCP Domains

Each of the 217 MCP protein sequences were manually searched against the Pfam (23) and SMART (in 'genomic mode') (21) databases for the ligand binding domain, transmembrane helix region, and HAMP linker domain annotations. We used both databases because one of the databases cannot annotate the ligand binding domain in many cases. Ligand binding domains annotated as TarH were recorded as four helix bundle sensory module for signal transduction (4HB) as TarH belongs to this domain (24). For transmembrane regions identification in the SMART database, the algorithm TMHMM2 (25) was used in conjunction with signal peptide prediction with the SignalP v4.0 program (26). Since we found a high number of false-negative annotations with both Pfam and SMART databases, we proceeded to identify the transmembrane helix regions with the DAS-TM filter algorithm on the DAS-TM filter server (27, 28). A score above 3.0 was used as a cut-off value for transmembrane helix region determination. Additionally, the transmembrane helix region that was determined as signal peptide by the SignalP v4.0 program was not counted (26). When false negative annotations on HAMP domain were found in both Pfam and SMART HAMP domain models, the presence of HAMP domain was recorded upon manual inspection of unidentified short region between the ligand binding domain and the signaling domain.

Determination of MCP classes and heptad nomenclature of the signaling domain

Based on the size of the ligand binding domain and the number of transmembrane regions, the 217 MCP were assigned to different topology classes as described previously (12,

29, 30). The genomes of the 14 bacterial strains were searched in the MiST database since it is the standard database that contains the information for the heptad nomenclature of the MCP signaling domain (31). We found that *Lonsdalea quercina* genomes in the MiST database do not have the strain name listed. For *Erwinia amylovora* and *Yersinia enterocolitica* subsp. *enterocolitica,* we could not find the type strain genome in the MiST database. In these cases, we chose the genome in the MiST database whose MCP protein sequences matched with the MCP protein sequences obtained from the type strain genome in the ASAP database (15, 16). Then, we used the heptad nomenclature as determined by the MiST database (31). We discovered that the number of MCP for *Brenneria salicis* ATCC 15712, *Hafnia alvei* ATCC 13337, and *Salmonella* Typhimurium LT2 in the MiST database was less than the number of MCP listed in the ASAP database, therefore, we were unable to determine the heptad nomenclature for these missing MCP.

RESULTS

Phylogeny of ABC transporter families

Classifications of ABC transporter families are often based on NBD sequences because of their conservation across transporters. Thus, we extracted 108 NBD sequences, including 106 in *P. atrosepticum* SCRI 1043 and two in *E. coli* K-12 MG1655, for phylogenetic analyses. The 108 NBD share a common ancestral root, indicating they likely form a super transporter group (Figure 2.1). The NBD can be classified into ten families, each with two or more members, most of which are predicted to be part of transporters that import or export sugars, peptides, amino acids, iron siderophores, lipids, or toxins (Table 2.2). The predicted substrates or functions for fourteen members are unknown, including transporters in family 3 (YbiT, YheS, Uup, YjjK),

family 5 (YdcT, ABL-0063625, ABL-0064587, ABL-0061599, ABL-0061724), family 6 (CydC, ABL-0064271), family 9 (ABL-0064453, ABL-0061768), and family 10 (YadG).

The NBDs are divided into 10 families based on *E. coli* transporters (8), so additional families may occur. The deep branches of families 4 and 5 suggest that these families could be subdivided. For example, the cluster containing ABL-0062415 may be a separate family; it contains only NBD from *P. atrosepticum* and their substrates are unknown. Within this clade, family 2, a group of peptide importers, is closely related to family 9, a group of exporters required for essential functions, including cell division and membrane elongation (FtsEX), outer-membrane integrity (LolCDE), and multidrug or enterotoxin resistance (MacAB). Putative transporters in family 3 show close relationship to family 6 exporters, which are involved in various multidrug efflux pumps and protease, lipid, toxin secretion machineries, indicating that members in family 3 may also play a role in exporting molecules. Unlike other families that usually form a clear subclade, NBDs in family 7 (YbhF and YbiH/YhiG) cluster with two members (LptB and YadG) in family 10 and are closely related to the family 1 sugar importers.

Conserved motifs of NBDs

The NBDs within an ABC transporter family often contain six conserved motifs, each with specific patterns of consensus amino acid residues, regardless of which family they are classified in. These six conserved structural elements are the Walker A motif, Q-loop, Walker B motif, H-motif, D-loop, and the signature motif, which is unique in ABC transporters (Table 2.5). The NBDs in family 3 and family 6 lack the Q-loop, which is usually between the Walker A and signature motifs. The Walker B motif is immediately preceded by the D-loop across all ABC families, and the NBD members in family 1 and family 3 seem to contain more than one signature motifs, Walker B, or D-loop regions in each domain. In addition to these shared

characteristics of NBD amino acid sequences, we also observed additional conserved regions in some ABC families. For example, NBD components belonging to family 2, a group of peptide importers, contain a motif with 26 amino acid residues (VE**********P*HPYT**LL***P) in the C-terminal region. This conserved motif is predicted to be present in oligopeptide/dipeptide transporters. In addition, one or more transmembrane regions are found in NBDs of family 6 and 7 and are upstream of the Walker A motif in family 6 or downstream of all other conserved motifs in family 7 (data not shown).

Diversity of ABC transporters in Enterobacteriales

After identifying orthologous proteins of each NBD component among the 70 bacteria, we examined the presence of these ABC members in the seven clades of Enterobacteriales (data not shown). The number of ATP-binding proteins varies among genera, with the exception of putative transporters in family 3, where all five NBDs were conserved across all genomes examined (data not shown). Peptide transporters, amino acid transporters, and iron-siderophore transporters are highly duplicated in Enterobacteriales compared to other transporter families (Table 2.2 & Figure 2.2). Despite the variation, one or more transporter systems within each family is conserved within each strain of the genera examined and a total of 26 transporters are conserved within the Enterobacteriales. These conserved systems include Rbs in family 1, Dpp and Sap in family 2, all members in family 3, Glt, Pst, Art and Met in family 4, Thi, Cys and Yrb in family 5, Cyd, Msb and Mdl in family 6, Znu and Hmu in family 8, Ybb, Lol and Fts in family 9, and Mod, Yhb/Lpt and Yad in family 10 (data not shown).

Bacteria that infect plants and, sometimes, those that infect animals, tend to encode more ABC transporters per genome than non-pathogenic bacteria (Figure 2.2). For example, soft-rot *Pectobacterium* and *Dickeya* possess higher numbers of sugar transporter systems than

commensal bacteria in genus *Hafnia* or a fresh-water dwelling bacterium *Budvicia aquatica* (data not shown). Similar patterns can be found between plant pathogens and non-pathogenic bacteria in peptide transporters, amino acid transporters, family 5 transporters, family 6 and 9 exporters, and iron-siderophore transporters. Animal pathogens in the genus of *Yersinia*, *Enterobacter* or *Escherichia* have greater numbers of transporters in families 1, 5, 8, and 10 than those in *Hafnia* and *Budvicia*. However, *M. morganii*, a bacterium that mainly infects immunocompromised people, encodes the least number of many ABC transporters. In addition, most strains of *Enterobacter*, *Pectobacterium*, *Dickeya*, *Brenneria*, or *Yersinia* encode the Ybh system but not the Yhi system (family 7) in their genomes. Yet neither of these transporters are found in *Erwinia* and *Lonsdalea* (Figure 2.2).

Domain structures of ABC transporter members in Enterobacteriales

In prokaryotes, ABC transporters are often comprised of single (NBD or TMD) and half structures (TMD-NBD or NBD-TMD). In Enterobacteriales, ABC transporter loci mainly encode an NBD-TMD half structure (42.6%), followed by an NBD-TMD-TMD structure (33.3%) and an NBD single structure (13.9%). Other structures such as NBD-NBD (13.9%) and TMD-TMD-NBD-NBD (6.5%) also occur (Figure 2.3A). Of the 26 conserved ABC transporters shared by bacteria in Enterobacteriales (Table 2.2 & Figures 2.4-2.13), most have an NBD single structure or half structure. For example, the Ybi, Yhe, Uup, Yjj, ModF, Msb, Ybb, and Yhb operons have single structure, accounting for approximately one third of the 26 conserved transporter systems. Another third of the shared transporters display the half structure, including Rbs, Met, Thi, Yrb, Znu, Hmu, Fts, ModC, and Yad. Despite these variations in the domain compositions, most of ABC importers also contain one or two substrate-binding proteins, located either upstream or downstream of an NBD or TMD (Figures 2.4-2.13). Several ABC operons also contain

additional subunits or genes, including enzymes, receptors, transcriptional regulators, and secretion proteins. These expanded operons are abundant in soft-rot Pectobacteriaceae compared to other Enterobacteriales. Thus, we attempt to describe these differences of the operon structures based on NBD classifications.

Family 1: ***Sugar transporter family*****.** Compared to other families, this set has relatively simple structures (Figure 2.3B & Figure 2.4). Six paralogs (Rbs, Ara, Xyl, ABL-0062988, ABL-0062716 and ABL-0062208) in this family are half structure (NMD-TMD) transporters, wherease YtfQRT, a galactofuranose transporter, contains duplicated permeases (NMD-TMD-TMD). RbsABC, a ribose importer, is found in all 70 strains, thus it is a conserved sugar transport system in Enterobacteriales. In addition, two importers (ABL-0061716 and ABL-0062208) are mainly present in soft rot pathogens *Pectobacterium* and *Dickeya* (data not shown). The third putative importer (ABL-0062988) is identified in both plant and animal pathogens but not in *Budvicia, Hafnia, and Morganella.*

Family 2: ***peptide transporter family.*** Eight of the 14 paralogs in this family have the NBD-TMD-TMD structure (YeiF, GsiA, ABL-0061245, ABL-0060660, ABL-0064758, ABL-0061245, ABL-0060339, and ABL-0063245), five have full structures (GntA, OppD, DppD, SapD, and ABL-0061805), while PhnK contains two NBDs (Figure 2.3B & Figure 2.5). In addition, Gnt, ABL-0061428, Abl-0060663, ABL-0061805, and ABL-0060339 are mainly present in the plant pathogens *Pectobacterium*, *Dickeya*, and/or *Brenneria*. Of these, the *gsi*-like operon (ABL-0064758), which likely imports glutathione like the Gsi system, is duplicated. However, unlike the *gsi* operon structure, this *gsi*-like operon also encodes three enzymes between the SBP (ABL-0064764) and the permeases (ABL-0064760 and ABL-0064759).

Family 3: putative transporters. Members of this family are putative ABC transporters with single structures, and their functions are mostly unknown. However, the ATP-binding protein ModF is adjacent to a family 10 molybdenum importer, ModABC, suggesting that transport of molybdenum may be mediated by ModABC, with ModF providing energy for the ModABC transporting system (Figure 2.6).

Family 4: amino acid transporters. These transporters comprise one of the largest ABC family in Enterobacteriales and have NBD-TMD structure or the half structure with duplicated TMD (Figure 2.3B & Figure 2.7). In most of the organisms, the *glt*, *pst*, *art* and *met* operons are highly conserved. The *pst* operon is unique among these, in that together with a regulator protein PhoU, it forms a complex to mediate activity of the two-component system, PhoRB, which regulates virulence factors and genes involved in assimilation of environmental phosphate. Other transporter operons, including ABL-0061090, ABL-0063255, ABL-0064815, ABL-0062415, ABL-0062324, and ABL-0064519, contain additional genes and are mainly found in the *Pectobacterium-Dickeya* clade.

Family 5: transporters with various substrates uptake. This is the largest ABC family in Enterobacteriales, consisting of 23 members responsible for uptake of diverse molecules, despite their conserved sequence similarity and operon structures (NBD-TMD or NBD-TMD-TMD structures; Figure 2.3B). Ten operons have extra genes, including Ssu, Mal, ABL-0062174, Nas, ABL-0061778, ABL-0064587, ABL-0060986, Ats, ABL-0064065, and Yrb (Figure 2.8). These additional genes are *tonB*, a *tonB* receptor, a *lacI* transcription regulator, and enzymes involved in metabolism. They may be actively associated with nutrient assimilation, energy transduction, regulation, and transport of certain substrates. ABC transporters, including Ydc, ABL-0062174, ABL-0063625, ABL-0061778, ABL-0060986, ABL-0061087, ABL-0064065, and ABL-

0061724 are mainly found in *Pectobacterium, Dickeya, Erwinia* or *Brenneria*. In addition, at least two importers in this family, MalKEFG and TogMNAB, play a role in bacterial chemotaxis, where the SBP may function as a co-sensor and interacts with its corresponding chemoreceptor (2-4, 11).

***Families 6, 7 and 9: mainly exporters*.** Most members in these families have NBD single structures or NBD-TMD structures (Figures 2.9, 2.10 & 2.12). These operons usually do not encode a substrate-binding protein due to their roles as exporters. Several efflux systems in family 6 are composed of an ABC exporter with NBD-TMD structure or a single structure with or without an outer membrane protein component (e.g. PrtF, HasF, and HylD). Six transporters (FusD, ABL-0063557, ABL-0061342, PrtD, ABL-0061741, and ABL-0064271) in family 6 are mainly found in the *Pectobacterium-Dickeya* clade. Of these six transporters, sequences and structures of the PrtDEF and HasDEF are highly homologous (Figure 2.9), and they play similar roles in secreting protein or protease. Similarly, two putative exporters in family 9, ABL-0064453 and ABL-0061768, are present in *Pectobacterium, Brenneria, Lonsdalea* or *Yersinia*. Moreover, the nucleotide-binding domain PhnL in family 9, along with an ATPase in family 2 (PhnK) and other enzymes, form a putative phosphonate importer (Figure 2.12).

Family 8 is involved in uptake of enzyme co-factors, such as of zinc, iron, and vitamin B12; however, only ZnuAVC (zinc uptake) and HmuSTUV (hemin uptake) are conserved across genera. There are at least six iron importers, suggesting there is a high competition for iron in the environment, and two (ABL-0062825 and ABL-0064591) of them are mainly found in the *Pectobacterium-Dickeya* clade (Figure 2.11). Interestingly, vitamin B12 importers are only encoded by a few genera (*Escherichia, Erwinia, Pectobacterium, Dickeya*, etc.), indicating these bacteria cannot synthesize the molecules and they must rely on these transporters for survival.

Family 10 is an ABC family with mixed transporters responsible for importing or exporting sugar, urea, branched amino acid or lipopolysaccharide. Members in this family have half or full structures and they shared a common ancestor with family 1 and family 7 (Figure 2.1). The only exporter system in this family is Yhb, which exports lipopolysaccharide, whereas the downstream SBP YhbN is a periplasmic chaperon that is involved in the LPT pathway (Figure 2.13).

Prevalence of chemotaxis systems and MCP in the order Enterobacteriales

The plant-associated Enterobacteriales have a higher number of *mcp* genes in their genomes than animal-associated bacteria, ranging from 11 to 41 *mcp* in the plant-associated genera, and from 0 to 11 *mcp* in the animal associated genera (Table 2.4). Among the 14 strains examined, only *Budvicia* lacks a chemotaxis system. The other 13 strains encode the core chemotaxis proteins CheA, CheW, CheY, CheR, CheB and CheZ. Additionally, *Hafnia, Yersinia, Dickeya, Erwinia, Enterobacter, Lonsdalea, P. atrosepticum* and *P. carotovorum* have an additional adaptor protein, CheV (Table 2.4). Interestingly, these strains also have 10 MCP or more in their genome.

MCP are divided into seven main classes (24H, 28H, 34H, 36H, 38H, 40H and 44H) and into five minor classes (42H, 48H, 52H, 58H, and 64H) based on the number of heptads in the cytoplasmic signaling domains (32). Most of the Enterobacteriales MCP are 36H (96%), while a few MCPs are 24H (0.4%), 28H (0.9%), 44H (0.4%), 52H (0.4%), or unknown (1.8%) (Table 2.4).

MCP have diverse ligand binding domains that allow them to sense a variety of extracellular and intracellular signals. MCP encoding 4HB, PAS, HBM, single Cache1, single Cache2, Cache3-2 fusion, Cache2, NIT, PAS3-PAS4, and FIST domains are encoded in the

Enterobacteriales. MCP with NIT, single Cache2, Cache3-2 fusion, and FIST were only present in the genomes from plant-associated bacteria, which encode both a greater number and a greater diversity of MCP (Table 2.3). Excitingly, the ligand binding domain of 10 MCP present in Enterobacteriaceae, Erwiniaceae, Hafniaceae, Morganellaceae, and Pectobacteriaceae do not match any characterized domain, suggesting that they encode novel domains. Four MCP in *Brenneria, Hafnia, Morganella,* and *Salmonella* appear to lack a ligand binding domain entirely.

MCP are categorized into four major classes (I-IV) based on membrane topology and into additional subclasses based on the ligand binding domain and transmembrane helices (5). Class I(a) cluster I is the most common type of MCP in bacteria (29), including in the Enterobacteriales (Table 2.3). We found that 61% of the MCP were in Class I(a) cluster I and additional 3% were in class I(a), but without a subcluster, due to having an unknown ligand binding domain (Table 2.3). Of the remaining MCP, the majority were in class I(a) cluster II (16%). Enterobacteriales genera also encode MCP in class I(b) (9%), class II (6%), class III(c) (1.4%), class IV(a) (1.8%), and class IV(b) (1.8%).

Phylogeny of MCPs

The cytoplasmic signaling domain is the most conserved domain of MCP, so it was used for phylogenic analysis of 14 strains representing each family in Enterobacteriales. A total of 217 signaling domain sequences were analyzed and we found that the 217 MCPs originate from the same ancestral root and expanded into 22 clades. MCPs from the soft rot genera (*Pectobacterium* and *Dickeya*) in the Pectobacteriaceae are present in all clades except Clade 7, which only comprised of a single MCP from *M. morganii* subsp. *morganii* (Table 2.3, Table 2.6 & Figure 2.14). Clades 3, 10, 13, and 21 include MCP from only Pectobacteriaceae. Therefore, the soft rot

Pectobacteriaceae not only have a larger number of MCP, their MCP are also more diverse than other Enterobacteriales.

There is not as much diversity in the other plant-associated Enterobacteriales as in the soft rot Pectobacteriaceae, suggesting that some aspects of being a soft rot pathogen require chemotaxis in response to a wider range of signals. MCP from the Enterobacteriaceae are present in 10 clades and MCP from Erwiniaceae are present in 7 clades. *Erwinia amylovora* infects woody plants, as does *Lonsdalea* and *Brenneria*, but there is little overlap in their MCP. Therefore, if these genera sense the same range of chemicals with MCP, they do this primarily through convergent evolution.

DISCUSSION

Comparison of ABC transporters and MCP in species within the Enterobacteriales order shows that the plant pathogenic Pectobacteriaceae are enriched in the number and diversity for ABC transporters and MCP compared to other Enterobacteriales. The Pectobacteriaceae family includes several devastating plant pathogens with worldwide distribution, including tree pathogens, such as *Lonsdalea* and *Brenneria*, and soft rot plant pathogens, such as *Pectobacterium* and *Dickeya*. It also includes *Sodalis*, an insect symbiont with a reduced genome (33). In addition to infecting many plant species, Pectobacteriaceae inhabits a diverse range of environments, including soil, water, insects, and clouds (34, 35).

Importing nutrients for cell metabolism, as well as exporting virulence proteins are crucial for bacterial survival and pathogenesis, and these selective transporters require robust transporting systems in bacterial genomes (36-38). The plant pathogens in Enterobacteriales encode a greater number of ABC transporters that are involved in the uptake of sugars (family 1), peptides (family 2), amino acids and derivatives (family 4 and 5), vitamins and iron-

siderophores (family 8), as well as in the export of drugs, toxins, protease, bacteriocins, lipoprotein and macrolide (family 6 and 9), which may confer competitive fitness advantage in nutrient acquisition. Plant-associated bacteria in this order also have an elevated number of MCP genes compared to animal-associated bacteria, which is congruent with other studies (39). Thus, we postulate that the high number of MCP genes in plant-associated bacteria may increase their fitness in fluctuating and diverse habitats (29), such as when the bacteria need to respond to wounds in plants, which are fluctuating and potentially hostile environment, but which are also an entry way into an environment that the pathogens can exploit (40).

We hypothesize that enrichment of MCP and ABC transporters may be linked in the plant pathogens since these two systems are sometimes functionally linked in bacteria. MCP help bacteria move towards chemicals that may be imported through ABC transporters and MCP may interact directly with the SBP of ABC transporters (2-4). For example, MalE, a maltose-substrate binding protein of the *E. coli* maltose transporter (MalEFGK in family 5), together with its bound maltose, interacts with the Tar chemoreceptor, resulting in cell taxis toward maltose (41). Another ABC importer in family 5, TogMNAB, which imports peptide oligomers and dimers, is associated with chemotaxis towards oligogalacturonides (2). In addition, some ABC transporters, such as FtsEX, TbpA-ThiPQ and LivFGHKM, are co-located with or adjacent to MCP in *Pectobacterium*. Therefore, the enrichment of ABC transporters and MCP could be coupled together and enrichment in these two systems may have been an important step in the evolution of plant pathogenesis.

The inferred phylogeny of ATP-binding proteins representing ten ABC families, mainly from the soft rot pathogen *P. atrosepticum* SCRI 1043, demonstrates that these NBDs have a common ancestral root (i.e. a superfamily). However, we cannot exclude the possibility that

some of ABC genes may be transferred from a more distantly related bacteria because no outgroup was included in the NBD phylogenetic tree (42-44). As expected, all NBDs examined in this study are identified with conserved motifs, including the Walker A (G**G*GKSTL), Signature motif (LSGGQ), Walker B (****DEPT), D-loop (D) and H motif (H) (Table 2.5). Interestingly, a transmembrane region was found in NBD of both family 6 and 7 exporters, suggesting an NBD and a TMD fused together during the evolution of these bacteria. A unique oligopeptide and dipeptide region was identified at C-terminal of family 2 transporters, indicating that this feature may be required for a functional interaction between NBD and the other domains of an ABC transporter for the uptake of peptides.

The transporters in family 4 and 5 are the most abundant families in Enterobacteriales, representing approximately 18% and 21% of the total ABC transport machinery, whereas family 7 is the least abundant family, accounting for 2% of the total ABC transporters (Figure 2.2). It is noteworthy that the NBDs in family 5 diverged from family 4, and most members in these two families have similar structures, including a half structure or a TMD-TMD-NBD/NBD-TMD-TMD organization (Figures 2.1-2.3). Interestingly, ABC transporters in family 5 are involved in the uptake of more diverse and complex substrates, some of which can be synthesized from amino acids or other substrates that may be imported by family 4 transporters.

The single structure NBDs of family 3 transporters branch from family 6 ATP-binding proteins, whose ABC operons display mostly single NBD and NBD-NBD structures (Figures 2.1 & 2.3). Every NBD of family 3 contains two sets of conserved motifs, including Walker A motif, Signature motif, Walker B motif, D-loop and H motif, but this pattern is not observed in members of family 6 (Figures 2.15 & 2.16). This suggests that NBDs in family 3 may originate by the fusion of two NBDs in family 6. Interestingly, no TMDs are found in ABC operons in

family 3, indicating NBD members in this family might complement energy needs for other biological processes (i.e. DNA repair, ribosome biogenesis) rather than being a functional transporter (45). Conversely, at least one transmembrane region is predicted in the NBD sequences in family 6 (Table 2.5). We postulate that the ancestral NBDs were fused to the TMD during their evolutionary processes, so they are able to export molecules to bacterial periplasmic space and actively participate in a type I secretion system (8-10, 46).

Multiple paralogs with similar functions are uniquely present in *Pectobacterium*, *Dickeya*, or *Brenneria* genomes, and these duplicated paralogs often have more complex operon structures than the conserved ABC transporters among different genera (Figures 2.4-2.13). Considering their broad-host range, the higher number of transporters may be necessary for soft rot pathogenesis since these pathogens colonize stems, roots, or tubers from numerous plant species and the nutrient composition and/or plant secondary metabolite profiles vary among hosts and tissues or dependent on environmental conditions (34, 38, 47-50). These results are in agreement with previous findings in *Pseudomonas syringae* pv. *tomato* and *Agrobacterium tumefaciens* (49, 51, 52), and hence provide further evidence that ABC transporters are used by phytobacterial pathogens to adapt to nutrient-diverse environments.

These nutrient-diverse environments also may explain the abundance of MCP in the soft rot Pectobacteriaceae. Chemotaxis plays an important role in bacterial fitness since it allows bacterial cells to integrate multiple signals and respond to chemical and energy gradients. Bacteria in the order Enterobacteriales appear to use chemotaxis mainly for regulating flagellar motility. *B. aquatica*, which is a water-associated bacterium, is the only species in our study that lacked a chemotaxis system, despite having flagellar apparatus and ability to be motile at certain temperatures (53). The other 13 bacterial strains possess proteins that are central to the

chemotaxis system, including CheA, CheW, CheY, CheR, CheB, CheZ, and MCP. Additionally, the strains that have 10 or more MCPs in their genome also encode an extra adaptor protein, CheV, which functions similarly to CheW, in linking CheA to MCP to form a stable protein complex (54). CheV was likely present in the common ancestor of the Enterobacteriales, but was lost in some genera, primarily the animal pathogens (54).

The 217 MCP appear to be inherited from a common ancestral root and further diversification occurred (Figure 2.14). The resulting phylogenetic tree separated the 217 MCP into 22 clades. MCP in some clades, namely Clade 1, 4, 5, 11, 14, 16, and 19, have undergone domain swapping events as MCP within these clades encode different ligand binding domains (Table 2.3 & 2.6). Plant-associated bacteria have more diversity in their MCP ligand binding domains, which supports the hypothesis that MCP assist in the adaptation of plant-associated bacteria to multiple and fluctuating environments. Four of the MCP clades identified (clade 3, 10, 13, and 21) are only found in plant pathogenic Pectobacteriaceae and the plant pathogen MCP had a greater diversity of ligand binding domains than non-plant pathogens. In addition, our analysis suggests that novel ligand-binding domains not previously described may be present in MCP from some of the species analyzed.

The MCP and ABC transporters of plant pathogens have been little examined, in part because the large number of these genes and potential duplication of function complicated analysis. Our studies and other bioinformatic analyses (7, 55) demonstrate the importance of MCP and ABC transporters to plant pathogenesis. These studies also aid in identification of proteins with potential novel functions, such as the novel ligand binding domains in the plant pathogen MCPs postulated in this work.

Table 2.1: List of bacterial strains in the order Enterobacteriales used in this study.

Clade	Family	Type Genus	Type Strain	Host of Isolation	References
Enterobacter-Escherichia	Enterobacteriaceae	*Enterobacter*	*Enterobacter asburiae* ATCC 35953	Human female lochia exudate	(56)
			Enterobacter bugandensis EB-247	Human newborn neonatal blood	(57)
			Enterobacter cloacae subsp. *cloacae* ATCC 13047	Human cerebrospinal fluid	(58)
			Enterobacter hormaechei subsp. *hormaechei* ATCC 49162	Human sputum	(59)
			Enterobacter hormaechei subsp. *oharae* EN-314 = DSM 16687	Human infant mouth swab	(59)
			Enterobacter hormaechei subsp. *steigerwaltii* EN-562 = DSM 16691	Human skin wound	(59)
			Enterobacter ludwigii EN-119	Human urine	(60)
			Enterobacter mori LMG 25706	Mulberry (*Morus alba* L.)	(61)
			Enterobacter roggenkampii DSM 16690	Human stool	(62, 63)
			Enterobacter soli ATCC BAA-2102 = LF7	Soil	(64)
		Escherichia	*Escherichia albertii* NBRC 107761 = Albert 19982	Human child diarrheal stool	(65)
			Escherichia coli ATCC 11775	Human urine	(66)
			Escherichia fergusonii ATCC 35469	Human stool	(67)
			Escherichia marmotae HT073016	Himalayan marmot stool (*Marmota himalayana*)	(68)
		Salmonella	*Salmonella* Typhimurium LT2	Unknown	(69)
Erwinia-Pantoea	Erwiniaceae	*Erwinia*	*Erwinia amylovora* ATCC 15580 = NCPPB 683	Pear (*Pyrus communis*)	(70)

			Erwinia billingiae NCPPB 661 = LMG 2613	Pear (*Pyrus communis*)	(71, 72)
			Erwinia gerundensis EM595	Pear leaves (*Pyrus communis*)	(73)
			Erwinia iniecta B120	Russian wheat aphid (*Diuraphis noxia*)	(74)
			Erwinia oleae DAPP-PG 531	Olive tree knots (*Olea europaea* L.)	(75)
			Erwinia persicina ATCC 35998	Tomato	(76, 77)
			Erwinia piriflorinigrans CFBP 5888	Pear necrotic blossoms	(78, 79)
			Erwinia pyrifoliae DSM 12163	Asian pear (*Pyrus pyrifolia* Nakai)	(80)
			Erwinia tasmaniensis Et1/99	Apple blossoms	(81)
			Erwinia teleogrylli SCU-B244	Chinese crickets (*Teleogryllus occipitalis*)	(82)
Pectobacterium-Dickeya	Pectobacteriaceae	*Pectobacterium*	*Pectobacterium atrosepticum* SCRI 1043[a]	Potato stem (*Solanum tuberosum*)	(83)
			Pectobacterium atrosepticum ATCC 33260 = CFBP 1526 = ICMP 1526 = LMG 2386 = NCPPB 549	Potato (*Solanum tuberosum*)	(84)
			Pectobacterium betavasculorum ATCC 43762 = CFBP 2122 = ICMP 4226 = LMG 2464 = NCPPB 2795	Sugar beet	(84)
			Pectobacterium carotovorum subsp. *brasiliense* 1692 (*Pectobacterium brasiliensis*)	Potato (*Solanum tuberosum*)	(35, 85)
			Pectobacterium carotovorum subsp. *carotovorum* ATCC 15713 = CFBP	Potato (*Solanum tuberosum*)	(84)

	2046 = ICMP 5702 = LMG 2404 = NCPPB 312		
	Pectobacterium carotovorum subsp. *odoriferum* NCPPB 3839 = CFBP 1878 = ICMP 11553 = LMG 5863 = NCPPB 3839	Chicory	(84, 86)
	Pectobacterium parmentieri RNS 08-42-1A = CFBP 8475	Potato (*Solanum tuberosum*)	(87, 88)
	Pectobacterium polaris NIBIO 1006	Potato tuber (*Solanum tuberosum*)	(89)
	Pectobacterium wasabiae ATCC 43316 = CFBP 3304 = ICMP 9121 = LMG 8404 = NCPPB 3701 = SR91	Japanese horseradish (*Eutrema wasabi*)	(84, 90)
Dickeya	*Dickeya aquatica* 174/2	River water	(91)
	Dickeya chrysanthemi ATCC 11663 = NCPPB 402	Florist's chrysanthemum (*Chrysanthemum morifolium*)	(92)
	Dickeya dadantii subsp. *dadantii* NCPPB 898	Pelargonium plant	(92, 93)
	Dickeya dadantii subsp. *dieffenbachiae* NCPPB 2976	*Dieffenbachia*	(92)
	Dickeya dianthicola NCPPB 453	*Dianthus*	(94)
	Dickeya fangzhongdai DSM 101947	Pear branches (*Pyrus pyrifolia* cv. 'Cuiguan')	(95)
	Dickeya paradisiaca ATCC 33242 = NCPPB 3532	Plantain	(92)
	Dickeya solani IPO 2222	Potato (*Solanum tuberosum*)	(94)
	Dickeya zeae NCPPB 2538	*Zea mays*	(96)

		Brenneria	*Brenneria goodwinii* FRB 141	*Quercus robur* (European oak)	(97)
			Brenneria roseae subsp. *americana* LMG 27715	*Quercus kelloggii* (California Black Oak)	(98)
			Brenneria salicis ATCC 15712 = DSM 30166 = LMG 2698	Willow (*Salix* spp.)	(99)
		Lonsdalea	*Lonsdalea britannica* LMG 26267	*Quercus robur* (European oak)	(100)
			Lonsdalea iberica LMG 26264	Mediterranean oak	(100)
			Lonsdalea populi LMG 27349	*Populus*× *euramericana* (*Populus canadensis* Moench)	(100)
			Lonsdalea quercina ATCC 29281 = CFCC 13731	Oak	(100)
		Sodalis	*Sodalis praecaptivus* ATCC BAA-2554 = DSM 27494 = HS	Human wound	(101, 102)
Yersinia-Serratia	Yersiniaceae	*Yersinia*	*Yersinia aldovae* ATCC 35236	Drinking water	(103)
			Yersinia aleksiciae strain 159 = Y159	Human stool	(104)
			Yersinia bercovieri ATCC 43970	Human stool	(105)
			Yersinia enterocolitica subsp. *enterocolitica* ATCC 9610 = biotype 1B/New World strain	Human cheek wound	(106, 107)
			Yersinia enterocolitica subsp. *palearctica* Y11 = Old World strain	Human stool	(107, 108)
			Yersinia entomophaga ATCC BAA-1678 = MH96	New Zealand grass grub larvae (*Costelytra zealandica*)	(109)

			Yersinia frederiksenii ATCC 33641 = CIP 80-29	Sewage	(110)
			Yersinia intermedia ATCC 29909	Human urine	(111)
			Yersinia kristensenii ATCC 33638 = CIP 80-30	Human urine	(112)
			Yersinia massiliensis CCUG 53443	Water, dialysis unit from hospital water distribution system	(113)
			Yersinia mollaretii ATCC 43969	Soil	(105)
			Yersinia nurmii CIP 110231 = APN3a-c	Broiler meat package	(114)
			Yersinia pekkanenii CIP 110230 = ÅYV7.1KOH2	Lettuce	(115, 116)
			Yersinia pestis ATCC 19428	Unknown	(117, 118)
			Yersinia pseudotuberculosis ATCC 29833	Turkey[b]	(117, 119)
			Yersinia rohdei ATCC 43380	Dog feces	(120)
			Yersinia ruckeri ATCC 29473	Rainbow trout (*Salmo gairdneri*)	(121)
Hafnia-Edwardsiella	Hafniaceae	*Hafnia*	*Hafnia alvei* ATCC 13337	Unknown	(122-124)
			Hafnia paralvei ATCC 29927	Human clinical specimen	(123)
Proteus-Xenorhabdus	Morganellaceae	*Morganella*	*Morganella morganii* subsp. *morganii* ATCC 25830 = NBRC 3848	Human diarrheal stool	(125-127)
Budvicia	Budviciaceae	*Budvicia*	*Budvicia aquatica* ATCC 35567[c] = DSM 5075	Well water	(53, 128)

Abbreviations:
ATCC, American Type Culture Collection, Manassas, Virginia, USA; **CCUG**, Culture Collection University of Gothenburg, Göteborg, Sweden; **CFBP**, CIRM-CFBP Collection Française de Bactéries Associées aux Plantes, Angers, France; **CIP**, Collection of the Institute Pasteur, Paris, France; **cv.**, cultivar; **ICMP**, International Collection of Microorganisms from Plants, Auckland, New

Zealand; **LMG**, BCCM/LMG Bacteria Collection, Laboratorium voor Microbiologie, Ghent, Belgium; **NCPPB**, National Collection of Plant Pathogenic Bacteria, Sand Hutton, York, England; **NIBIO**, Norwegian Institute of Bioeconomy Research, As, Norway.

[a]Not a type strain.

[b]Information retrieved from https://www.dsmz.de/collection/catalogue/details/culture/DSM-8992.

[c]The type strain number on ASAP database differs from the literature (ATCC 25567).

Table 2.2: The NBDs used for ATP-binding protein orthologs identification and phylogeny analysis.

ABC Family	NBD Component	Accession NO.	GenBank NO.	Transporter	Substrate	Function
1	RbsA*	ABL-0060210	CAG72935.1	RbsABC	D-ribose	Monosaccharide importer
	AraG	ABL-0062536	CAG75175.1	AraFGH	L-arabinose, fructose, xylose	Monosaccharide importer
	XylG	ABL-0060300	CAG73019.1	XylFGH	D-xylose, ribose	Monosaccharide importer
	YtfR	ABL-0064566	CAG77132.1	YtfQRT-YjfF	D-galactofuranose	Galactofuranose importer
	-	ABL-0062988	CAG75619.1	-	-	Putative sugar importer
	-	ABL-0061716	CAG74370.1	-	-	Putative sugar importer
	-	ABL-0062208	CAG74850.1	-	-	Putative sugar importer
2	GntA	ABL-0064398	CAG76970.1	GntABCMN	Peptides	Peptide importer
	-	ABL-0061245	CAG73916.1	-	Glutathione	Oligopeptide importer
	-	ABL-0060660	CAG73360.1	-	Oligopeptide	Putative oligopeptide importer
	GsiA like	ABL-0064758	CAG77316.1	-	Glutathione	Oligopeptide importer
	-	ABL-0061805	CAG74458.1	-	-	ABC Importer
	-	ABL-0060274	CAG72993.1	-	Oligopeptide	Oligopeptide importer
	OppD	ABL-0062585	CAG75224.1	OppABCDF	Oligopeptide	Oligopeptide importer
	DppD*	ABL-0064729	CAG77287.1	DppABCDF	Dipeptide, 5-aminolevulinic acid (ALA)	Dipeptide importer
	SapD*	ABL-0062234	CAG74876.1	SapABCDF	Cationic peptide	Oligopeptide importer
	YejF	ABL-0063008	CAG75639.1	YejABEF	Probable Oligopeptide	Putative importer
	GsiA	ABL-0063108	CAG75735.1	GsiABCD	Glutathione	Oligopeptide importer
	-	ABL-0060339	CAG73058.1	-	-	Peptide importer
	-	ABL-0063245	CAG75860.1	-	-	Peptide importer
	PhnK	ABL-0060710	CAG73409.1	PhnKL(##)	-	Putative phosphonate transporter
	YbiT*	ABL-0063295	CAG75909.1	#	-	Putative transporter

	YheS*	ABL-0064383	CAG76955.1	#	-	Putative transporter
3	Uup*	ABL-0062798	CAG75433.1	#	-	Putative transporter
	YjjK*	ABL-0064218	CAG76799.1	#	-	Putative transporter
	ModF*	ABL-0061644	CAG74300.1	#	-	Putative molybdenum transporter
	GlnQ	ABL-0063043	CAG75670.1	GlnHPQ	Glutamine	Amino acid importer
	-	ABL-0063853	CAG76437.1	-	-	Amino acid importer
	YhdZ	ABL-0060460	CAG73162.1	YhdWXYZ	-	Putative amino acid importer
	HisP	ABL-0063334	CAG75946.1	HisJMPQ	L-lysine, L-arginine, L-ornithine	Amino acid importer
	GltL*	ABL-0061557	CAG74221.1	GltIJKL	Glutamate/aspartate	Amino acid importer
	PstB*	ABL-0064811	CAG77369.1	PstSABC-phoU	Phosphate	High-affinity phosphate importer
	-	ABL-0063800	CAG76390.1	-	-	Putative amino acid importer
	-	ABL-0063449	CAG76059.1	PstABC	Phosphate	Phosphate importer
4	-	ABL-0062956	CAG75588.1	-	Methionine	Methionine importer
	ArtP*	ABL-0062935	CAG75567.1	ArtIMQP/ArtJMQP	L-arginine	Amino acid importer
	MetN*	ABL-0063838	CAG76422.1	MetINQ	D-methionine	Methionine importer
	-	ABL-0061090	CAG73767.1	-	Arginine, ectoine, hydroxyectoine	Amino acid importer
	OccP	ABL-0064409	CAG76981.1	OccJMPQ	Octopine	Amino acid importer
	-	ABL-0063255	CAG75870.1	-	-	Amino acid transporter
	-	ABL-0064815	CAG77373.1	-	-	Amino acid importer
	-	ABL-0062415	CAG75055.1	-	-	Amino acid importer
	MetN1	ABL-0062324	CAG74965.1	-	Methionine	Methionine importer
	-	ABL-0064519	CAG77090.1	-	-	Amino acid transporter
	ProV	ABL-0063810	CAG76400.1	ProVWX	Glycine betaine, L-proline	Glycine betaine/L-proline importer
	ThiQ*	ABL-0064163	CAG76744.1	TbpA-ThiPQ	Thiamine	Thiamine importer

AfuC/FbpC 1	ABL-0061748	CAG74401.1	FbpABC	Fe^{3+}	Fe^{3+} ion importer
SsuB	ABL-0064571	CAG77309.1	SsuABC	Aliphatic sulfonates	Aliphatic sulfonates importer
PotG	ABL-0062944	CAG75576.1	PotFGHI	Putrescine	Putrescine importer
CysA*	ABL-0063591	CAG76194.1	CysAWUP	Sulfate, thiosulfate	Sulfate/thiosulfate importer
PotA	ABL-0062713	CAG75351.1	PotABCD	Spermidine, putrescine	Spermidine/putrescine importer
MalK	ABL-0063462	CAG76072.1	MalEFGK	Maltose, maltodextrin	Maltose/maltodextrin importer
YdcT	ABL-0062312	CAG74953.1	#	-	ABC Importer
TogA	ABL-0062670	CAG75308.1	TogMNAB	Pectic oligomers or dimers	Pectic oligomers/dimers importer
SfuC	ABL-0063428	CAG76038.1	SfuABC	Fe^{3+}	Fe^{3+} ion importer
-	ABL-0062174	CAG74817.1	-	-	ABC Importer
NasD	ABL-0063278	CAG75892.1	NasDEF	Nitrate	Nitrate importer
-	ABL-0063625	CAG75880.1	-	-	Putative transporter
-	ABL-0061778	CAG74431.1	-	Nitrate, sulfonate, bicarbonate	Sulfonate importer
-	ABL-0064587	CAG77147.1	-	-	ABC Importer
-	ABL-0060986	CAG73665.1	-	-	Putative sugar importer
AtsC	ABL-0064634	CAG77193.1	-	Sulfate	Sulfate ester importer
-	ABL-0061087	CAG73764.1	-	-	Putative sugar importer
-	ABL-0064065	CAG76647.1	-	-	Putative sugar importer
-	ABL-0061599	CAG74257.1	-	-	ABC importer
-	ABL-0061724	CAG74378.1	-	-	ABC importer
TauB	ABL-0061783	CAG74436.1	TauABC	Taurine	Putative taurine importer

	YrbF/MlaF*	ABL-0060513	CAG73215.1	YrbCDEF/MlaBDEF	Phospholipid, cholesterol, gamma-HCH	Predicted toluene importer
	CydC*	ABL-0062919	CAG75551.1	##	-	Putative transporter
	MsbA*	ABL-0062821	CAG75456.1	MsbA(#)	Lipid A	Flip lipid A to the periplasmic side of the inner membrane; MsbA functions as NBD and TMD
	MdlB*	ABL-0061404	CAG74069.1	MdlAB(##)	-	Multidrug exporter; MdlB functions as NBD and TMD
6	FusD	ABL-0061115	CAG73792.1	FusCD(#)	-	Iron-sulphur importer; FusD functions as NBD and TMD
	MdlB4	ABL-0062679	CAG75317.1	#	-	Putative multidrug exporter
	-	ABL-0063557	CAG76166.1	#	-	Putative toxin secretion
	-	ABL-0061342	CAG74007.1	-	-	Putative toxin secretion
	PrtD	ABL-0063056	CAG75683.1	PrtDE	Protease	Protease secretion
	HasD	ABL-0061789	CAG74442.1	HasDE	Protein	Type I secretion
	-	ABL-0061741	CAG74395.1	#	-	Putative peptide exporter; functions as NBD and TMD
	-	ABL-0061346	CAG74011.1	#	-	Putative peptide exporter; functions as NBD and TMD
	-	ABL-0064271	CAG76851.1	-	-	ABC Importer
	UgpC	ABL-0064657	CAG77216.1	UgpABCE	*sn*-glycerol3-phosphate, glycerophosphocholine	*sn*-glycerol-3-phosphate importer

7[$]	YhiH/YhiG	ABE-0011381	BAE77807.1	YhiHI-YhhJ	-	Protein-chain elongation; in release of deacyl-tRNA
	YbhF	ABE-0002707	BAA35454.2	YbhFSR	Cefoperazone	Putative multidrug exporter
8	ZnuC*	ABL-0062749	CAG75385.1	ZnuABC	Zinc	High-affinity zinc importer
	YfeB	ABL-0062658	CAG75296.1	YfeABCD	Chelated iron	Chelated iron importer
	YiuC	ABL-0063532	CAG76141.1	-	Iron	Putative ferric-enterobactin importer
	BtuD	ABL-0062092	CAG74737.1	BtuCDE	Vitamin B12	Vitamin B12 importer
	CbrD	ABL-0063079	CAG75706.1	CbrABCD	-	Achromobactin importer
	FhuC	ABL-0063607	CAG76210.1	FhuBCD	Fe^{3+}-hydroxamate	Iron-hydroxamate importer
	FecE	ABL-0061319	CAG73984.1	FecABCDE	Fe^{3+}-dicitrate	Fe^{3+} dicitrate importer
	HmuV*	ABL-0062101	CAG74746.1	HmuSTUV	Hemin	Hemin importer
	-	ABL-0062825	CAG75460.1	-	-	Putative iron importer
	-	ABL-0064591	CAG77151.1	-	-	Putative Fe^{3+} importer
	-	ABL-0064073	CAG76655.1	-	-	Putative metal importer
	-	ABL-0063382	CAG75993.1	-	-	Putative importer
	CcmA	ABL-0062141	CAG74785.1	-	-	Cytochrome C biogenesis; heme exporter
9	YbbA*	ABL-0061469	CAG74133.1	YbbA(#)	-	Putative metal exporter
	-	ABL-0064453	CAG77025.1	-	-	ABC Importer
	PhnL	ABL-0060711	CAG73410.1	PhnCDEFGHIJKL MNOP	Phosphonate	Putative phosphonate transporter
	MacB	ABL-0061122	CAG73799.1	MacAB	14- and 15-membered lactones	Macrolide exporter
	LolD*	ABL-0062082	CAG74727.1	LolCDE	Lipoproteins	Lipoprotein exporter
	-	ABL-0061768	CAG74421.1	-	-	ABC exporter
	FtsE*	ABL-0064683	CAG77242.1	FtsEX	Amidase	Cell division

	LivG	ABL-0064676	CAG77235.1	LivFGHKM	Leucine, isoleucine, valine	High-affinity branched-chain amino acid importer
	ModC*	ABL-0061649	CAG74305.1	ModABC	Molybdenum	Molybdenum importer
10	UrtD	ABL-0062409	CAG75049.1	UrtDE	-	Putative urea importer
	YhbG/LptB*	ABL-0060507	CAG73209.1	#	Lipopolysaccharide	Lipopolysaccharide exporter
	YadG*	ABL-0063620	CAG76223.1	YadGH	-	ABC importer

-: not being annotated or reported.

*: represent conserved ABC transporters in Enterobacteriales.

#: represent an NBD.

$: *Escherichia coli* K-12 MG1655 as reference NBDs.

Table 2.3: Clade classification based on phylogeny analysis of the MCP cytoplasmic signaling domain in the order Enterobacteriales.

Clade	Species	ASAP Feature ID	Heptad Nomenclature of the Signaling Domain	Ligand Binding Domain (LBD)	Putative Localization of the Ligand Binding Domain	Number of Transmembrane Helix Region	HAMP Presence	MCP Class
1	*E. cloacae* subsp. *cloacae*	AKI-0001960	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	M. morganii subsp. *morganii*	BJA-0000045	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	E. amylovora	AYK-0000069	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	D. chrysanthemi	ATU-0003627	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
2	*P. atrosepticum*	ABL-0062322	36H	4HB	Periplasm	1	Yes	Class I(b)
	P. carotovorum subsp. *carotovorum*	BJR-0002238	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. amylovora	AYK-0003447	36H	Unknown	Periplasm	1	Yes	Class I(b)
	H. alvei	AMD-0000151	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	H. alvei	AMD-0000702	Unknown	Unknown	Periplasm	1	Yes	Class I(b)

	E. amylovora	AYK-0003184	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0003395	36H	Unknown	Periplasm	2	Yes[a]	Class I(a)[b]
3	*D. chrysanthemi*	ATU-0003658	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. atrosepticum	ABL-0060651	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0000473	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. atrosepticum	ABL-0063534	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0002977	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0001943	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0002628	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
4	*D. chrysanthemi*	ATU-0003063	36H	Cache2	Periplasm	2	Yes[a]	Class I(a) Cluster I
	Y. enterocolitica subsp. *enterocolitica*	BPH-0002254	36H	PAS	Cytoplasm	0	Yes	Class IV(a)
	L. quercina	BIZ-0000399	36H	PAS	Cytoplasm	2	Yes	Class II

	M. morganii subsp. *morganii*	BJA-0000014	36H	PAS	Cytoplasm	2	Yes	Class II
	D. chrysanthemi	ATU-0000843	36H	PAS	Periplasm	1	Yes	Class I(b)
	P. atrosepticum	ABL-0063893	36H	PAS	Cytoplasm	2	Yes	Class II
	P. carotovorum subsp. *carotovorum*	BJR-0003338	36H	PAS	Cytoplasm	2	Yes	Class II
5	*P. atrosepticum*	ABL-0063957	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	Y. enterocolitica subsp. *enterocolitica*	BPH-0004021	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0062794	36H	Cache2	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0002643	36H	Cache2	Periplasm	2	Yes[a]	Class I(a) Cluster I
	D. chrysanthemi	ATU-0003716	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0060608	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0000418	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0001467	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I

	D. chrysanthemi	ATU-0001412	36H	4HB	Cytoplasm	0	Yes	Class IV(a)
6	*H. alvei*	AMD-0000584	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	Y. enterocolitica subsp. *enterocolitica*	BPH-0000764	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0063906	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0003351	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	B. salicis	AMK-0003262	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0000830	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	L. quercina	BIZ-0000386	36H	4HB	Periplasm	1	Yes	Class I(b)
	E. cloacae subsp. *cloacae*	AKI-0000652	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	E. amylovora	AYK-0000468	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0000802	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I

	L. quercina	BIZ-0000364	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	B. salicis	AMK-0003230	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0063938	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0003386	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
7	*M. morganii* subsp. *morganii*	BJA-0003445	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
8	*D. chrysanthemi*	ATU-0002418	36H	PAS	Cytoplasm	2	Yes[a]	Class II
	E. amylovora	AYK-0000397	36H	PAS	Cytoplasm	2	Yes[a]	Class II
	H. alvei	AMD-0001112	36H	PAS	Cytoplasm	2	Yes[a]	Class II
	E. cloacae subsp. *cloacae*	AKI-0004512	36H	PAS	Cytoplasm	2	Yes[a]	Class II
	E. coli	ABE-0010087	36H	PAS	Periplasm	1	Yes[a]	Class I(b)
	S. Typhimurium	ABU-0100002	36H	PAS	Periplasm	1	Yes[a]	Class I(b)
9	*E. cloacae* subsp. *cloacae*	AKI-0001637	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	H. alvei	AMD-0002898	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II

	Y. enterocolitica subsp. *enterocolitica*	BPH-0002563	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	D. chrysanthemi	ATU-0003144	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	L. quercina	BIZ-0001695	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	B. salicis	AMK-0003355	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0061350	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	P. carotovorum subsp. *carotovorum*	BJR-0001071	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
10	*D. chrysanthemi*	ATU-0004009	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	L. quercina	BIZ-0002527	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	D. chrysanthemi	ATU-0004006	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I

	D. chrysanthemi	ATU-0004007	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. atrosepticum	ABL-0064672	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0004042	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. atrosepticum	ABL-0064673	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0004043	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	B. salicis	AMK-0001829	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	D. chrysanthemi	ATU-0004005	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. atrosepticum	ABL-0064445	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0003816	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0004040	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0004041	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I

	B. salicis	AMK-0001830	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. atrosepticum	ABL-0064670	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. atrosepticum	ABL-0064671	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
11	*D. chrysanthemi*	ATU-0002120	52H	FIST-FIST_C	Cytoplasm	0	Yes[a]	Class IV(a)
	Y. enterocolitica subsp. *enterocolitica*	BPH-0000227	24H	PAS3-PAS4	Cytoplasm	0	No	Class IV(a)
	S. Typhimurium	ABU-0099776	44H	No LBD	Cytoplasm	0	No	Class IV(b)
	H. alvei	AMD-0002795	28H	No LBD	Cytoplasm	0	No	Class IV(b)
	M. morganii subsp. *morganii*	BJA-0000488	28H	No LBD	Cytoplasm	0	No	Class IV(b)
	D. chrysanthemi	ATU-0000485	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	L. quercina	BIZ-0001961	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
12	*M. morganii* subsp. *morganii*	BJA-0000015	36H	Unknown	Periplasm	2	Yes	Class I(a)[b]
	E. cloacae subsp. *cloacae*	AKI-0001855	36H	Unknown	Periplasm	2	Yes	Class I(a)[b]
	S. Typhimurium	ABU-0095413	36H	Unknown	Periplasm	2	Yes	Class I(a)[b]

	Y. enterocolitica subsp. *enterocolitica*	BPH-0003177	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. cloacae subsp. *cloacae*	AKI-0002017	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0002023	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	B. salicis	AMK-0002809	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0062581	36H	4HB	Cytoplasm	3	Yes	Class III(c)
	P. carotovorum subsp. *carotovorum*	BJR-0002451	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
13	*B. salicis*	AMK-0001440	Unknown	No LBD	Cytoplasm	0	No	Class IV(b)
	P. atrosepticum	ABL-0062261	36H	PAS	Cytoplasm	2	Yes[a]	Class II
	P. carotovorum subsp. *carotovorum*	BJR-0002181	36H	PAS	Cytoplasm	2	Yes[a]	Class II
	D. chrysanthemi	ATU-0002449	36H	PAS	Cytoplasm	2	Yes[a]	Class II
	P. atrosepticum	ABL-0062262	36H	PAS	Cytoplasm	2	Yes	Class II
	P. carotovorum subsp. *carotovorum*	BJR-0002182	36H	PAS	Cytoplasm	2	Yes	Class II
14	*H. alvei*	AMD-0003114	36H	Unknown	Cytoplasm	6	Yes[a]	Class III(c)
	D. chrysanthemi	ATU-0000685	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I

	E. cloacae subsp. *cloacae*	AKI-0000598	36H	Cache3-2 Fusion	Periplasm	2	Yes	Class I(a) Cluster II
	E. amylovora	AYK-0000484	36H	Cache3-2 Fusion	Periplasm	1	Yes	Class I(b)
	D. chrysanthemi	ATU-0000810	36H	Cache3-2 Fusion	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0063929	36H	Cache3-2 Fusion	Periplasm	2	Yes	Class I(a) Cluster II
	P. carotovorum subsp. *carotovorum*	BJR-0003375	36H	Cache3-2 Fusion	Periplasm	2	Yes	Class I(a) Cluster II
	D. chrysanthemi	ATU-0001206	36H	NIT	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0060649	36H	NIT	Periplasm	2	Yes	Class I(a) Cluster II
	P. carotovorum subsp. *carotovorum*	BJR-0000471	36H	NIT	Periplasm	2	Yes	Class I(a) Cluster II
15	*M. morganii* subsp. *morganii*	BJA-0000631	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I

	E. cloacae subsp. *cloacae*	AKI-0004511	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	S. Typhimurium	ABU-0099999	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0002622	36H	4HB	Cytoplasm	3	Yes	Class III(c)
	B. salicis	AMK-0000941	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0062033	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0001966	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	L. quercina	BIZ-0000332	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0060384	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0000184	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0060383	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0000183	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I

16	*H. alvei*	AMD-0003657	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0002117	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0060281	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	P. carotovorum subsp. *carotovorum*	BJR-0000077	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0062843	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	P. carotovorum subsp. *carotovorum*	BJR-0002696	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	D. chrysanthemi	ATU-0001892	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0062844	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	P. carotovorum subsp. *carotovorum*	BJR-0002697	36H	HBM	Periplasm	2	Yes	Class I(a)

								Cluster II
17	*D. chrysanthemi*	ATU-0001813	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0001814	36H	4HB	Periplasm	1	Yes	Class I(b)
	E. amylovora	AYK-0003419	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	M. morganii subsp. *morganii*	BJA-0002462	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. cloacae subsp. *cloacae*	AKI-0004376	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	S. Typhimurium	ABU-0099818	36H	Unknown	Periplasm	2	Yes	Class I(a)[b]
	Y. enterocolitica subsp. *enterocolitica*	BPH-0000094	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0002967	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	L. quercina	BIZ-0002050	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	B. salicis	AMK-0003540	36H	Unknown	Periplasm	2	Yes	Class I(a)[b]
	P. atrosepticum	ABL-0061584	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I

	P. carotovorum subsp. *carotovorum*	BJR-0001264	36H	Unknown	Periplasm	2	Yes[a]	Class I(a)[b]
18	*D. chrysanthemi*	ATU-0002686	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0061941	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0001874	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. coli	ABE-0004744	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. cloacae subsp. *cloacae*	AKI-0002084	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	S. Typhimurium	ABU-0095332	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0000571	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0064219	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0003591	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
19	*D. chrysanthemi*	ATU-0002384	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II

	D. chrysanthemi	ATU-0002349	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	Y. enterocolitica subsp. *enterocolitica*	BPH-0003971	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	D. chrysanthemi	ATU-0000806	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0003382	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0000135	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	L. quercina	BIZ-0002810	36H	HBM	Periplasm	1	Yes	Class I(b)
	B. salicis	AMK-0001541	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0060292	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II
	P. carotovorum subsp. *carotovorum*	BJR-0000091	36H	HBM	Periplasm	2	Yes	Class I(a) Cluster II

20	*P. atrosepticum*	ABL-0064154	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0003529	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	M. morganii subsp. *morganii*	BJA-0002284	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
21	*P. carotovorum* subsp. *carotovorum*	BJR-0004142	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II
	P. atrosepticum	ABL-0062489	36H	Cache1	Periplasm	1	Yes[a]	Class I(b)
	D. chrysanthemi	ATU-0004051	36H	Cache1	Periplasm	2	Yes[a]	Class I(a) Cluster II
	P. carotovorum subsp. *carotovorum*	BJR-0002378	36H	Cache1	Periplasm	2	Yes[a]	Class I(a) Cluster II
22	*M. morganii* subsp. *morganii*	BJA-0002283	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. amylovora	AYK-0001843	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	S. praecaptivus	BOS-0000622	36H	4HB	Periplasm	1	Yes	Class I(b)
	H. alvei	AMD-0002493	36H	4HB	Periplasm	1	Yes	Class I(b)

	Y. enterocolitica subsp. *enterocolitica*	BPH-0001466	36H	4HB	Periplasm	1	Yes	Class I(b)
	E. cloacae subsp. *cloacae*	AKI-0001443	36H	4HB	Periplasm	1	Yes	Class I(b)
	E. coli	ABE-0006288	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	S. Typhimurium	ABU-0101078	Unknown	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. cloacae subsp. *cloacae*	AKI-0002276	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0061527	36H	4HB	Periplasm	1	Yes	Class I(b)
	P. carotovorum subsp. *carotovorum*	BJR-0001216	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. cloacae subsp. *cloacae*	AKI-0000620	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. cloacae subsp. *cloacae*	AKI-0001442	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. coli	ABE-0006290	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	S. Typhimurium	ABU-0096205	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	S. praecaptivus	BOS-0000621	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I

	D. chrysanthemi	ATU-0002417	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. amylovora	AYK-0002642	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. amylovora	AYK-0001846	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	Y. enterocolitica subsp. *enterocolitica*	BPH-0001465	36H	4HB	Periplasm	1	Yes	Class I(b)
	H. alvei	AMD-0002492	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	S. Typhimurium	ABU-0103914	Unknown	4HB	Periplasm	1	Yes	Class I(b)
	E. cloacae subsp. *cloacae*	AKI-0000777	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. coli	ABE-0014282	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	L. quercina	BIZ-0001845	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0002679	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	B. salicis	AMK-0001023	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I

	P. atrosepticum	ABL-0061949	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0001882	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	Y. enterocolitica subsp. *enterocolitica*	BPH-0003384	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	L. quercina	BIZ-0003081	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	B. salicis	AMK-0002535	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	D. chrysanthemi	ATU-0000514	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0001812	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. cloacae subsp. *cloacae*	AKI-0003301	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	E. amylovora	AYK-0002399	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	B. salicis	AMK-0003186	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	L. quercina	BIZ-0002289	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I

	B. salicis	AMK-0003715	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. atrosepticum	ABL-0061762	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0001432	36H	4HB	Periplasm	1	Yes[a]	Class I(b)
	L. quercina	BIZ-0002288	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	D. chrysanthemi	ATU-0001815	36H	4HB	Periplasm	2	Yes[a]	Class I(a) Cluster I
	B. salicis	AMK-0000801	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. atrosepticum	ABL-0062981	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I
	P. carotovorum subsp. *carotovorum*	BJR-0002775	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I

[a]HAMP presence was identified by manual inspection of unidentified short region between the ligand binding domain (LBD) and the cytoplasmic signaling domain (SD) sequences.

[b]Subcluster could not be determined due to unknown ligand binding domain.

Table 2.4: The number of chemotaxis system presents in the 14 strains of the order Enterobacteriales based on MiST database.

Family	Type Strain	MCP[b]	CheA	CheW	CheY	CheR	CheB	CheV	CheZ
Animal-associated bacteria									
Budviciaceae	*Budvicia aquatica* ATCC 35567	0	0	0	0	0	0	0	0
Enterobacteriaceae	*Escherichia coli* K-12 substr. MG1655	5 (36H)	1	1	1	1	1	0	1
Enterobacteriaceae	*Salmonella* Typhimurium LT2	6 (36H) 1 (44H) 2 (Unknown)	1	1	1	1	1	0	1
Hafniaceae	*Hafnia alvei* ATCC 13337	1 (28H) 8 (36H) 1 (Unknown)	1	1	1	1	1	1	1
Morganellaceae	*Morganella morganii* subsp. *morganii* ATCC 25830	1 (28H) 8 (36H)	1	1	1	1	1	0	1
Pectobacteriaceae	*Sodalis praecaptivus* ATCC BAA-2554 = HS1	2 (36H)	1	1	1	1	1	0	1
Yersiniaceae	*Yersinia enterocolitica* subsp. *enterocolitica*	1 (24H) 10 (36H)	1	1	1	1	1	1	1
Plant-associated bacteria									
Enterobacteriaceae	*Enterobacter cloacae* subsp. *cloacae* ATCC 13047 1	16 (36H)	2	2	2	2	2	1	2
Erwiniaceae	*Erwinia amylovora* ATCC 49946 (not type strain)	11 (36H)	2	2	2	2	2	1	2
Pectobacteriaceae	*Brenneria salicis* ATCC 15712	14 (36H) 1 (Unknown)	1	1	1	1	1	0	1
Pectobacteriaceae	*Dickeya chrysanthemi* ATCC 11663	40 (36H) 1 (52H)	1	1	1	1	1	1	1
Pectobacteriaceae	*Lonsdalea quercina*[a]	13 (36H)	1	1	1	1	1	1	1

Pectobacteriaceae	*Pectobacterium atrosepticum* SCRI 1043	36 (36H)	1	1	1	1	1	1	1
Pectobacteriaceae	*Pectobacterium carotovorum* subsp. *carotovorum* ATCC 15713	39 (36H)	1	1	1	1	1	1	1

[a]no strain listed in MiST database.

[b]the number of heptad conservation in MCP signaling domain is indicated in parenthesis.

Table 2.5: Conserved motifs among the NBDs of ten ABC transporter families in Enterobacteriales.

	Walker A	**Q-loop**	**Walker B**	**H-motif**	**Signature motif**	**D-loop**	**Oligopeptide/ dipeptide transporter, C-terminal region**	**Transmembrane regions**
Family 1	GENGAGKSTL	*I*Q	LILDEPT	Y(I/V)SHRL	LS*GQQQ*VE	A*L(S/T/D)	-	-
Family 2	GESGSGKS	*IFQ	LLIADEPT	ITH	LSGG**QRV*IA	TALD	VE*********P*HPYT**LL***P	-
Family 3	G*NG*GKSTL	-	LLLDEPT	VSH	LSGGE	N*LD	-	-
Family 4	G*SG*GKSTL	MVFQ	L*DEPT	*TH	LSGGQ*QRV*IARALA	SALD	-	-
Family 5	G*SG*GKST*L	**FQ	*L*DEP	VTH	*SGGQ*QR*A	**LD	-	-
Family 6	G**G*GKSTL	-	LDE	*TH	LSGGQRQ**A*	***D	-	Yes
Family 7	GPDG*GK(S/T)	*MPQ	*L*LDEP	THFM	LSGGM	(T/V)GVD	-	Yes
Family 8	G*NG*GKSTLL	***Q	**LDEP	V*HD*	LSGGQ*QRV	**LD	-	-
Family 9	G*SGSGKSTL	**FQ	LADEPT	VTH	LSGGEQQRV*	GNLD	-	-
Family 10	GPNGAGKTT	**FQ	LLLLDEP	L***H	LSGG*KRRLE	AG*D	-	-

*: indicate more than 3 amino acids were observed.

Table 2.6: Clade summary based on phylogeny analysis of the MCP cytoplasmic signaling domain in the order Enterobacteriales.

Clade	Heptad Nomenclature of the Signaling Domain	Ligand Binding Domain[a]	Putative localization of the ligand binding domain	Number of transmembrane helix region	HAMP Presence	MCP Class[b]	Animal-associated genera	Plant-associated genera
1	36H	4HB, Cache1	Periplasm	2	Yes[c]	Class I(a) Cluster I and Cluster II	*Morganella*	*Dickeya, Erwinia, Enterobacter*
2	36H, unknown	4HB, unknown	Periplasm	1, 2	Yes[c]	Class I(a)[d], Class I(a) Cluster I, Class I(b)	*Hafnia*	*Erwinia, Pectobacterium*
3	36H	4HB	Periplasm	2	Yes[c]	Class I(a) Cluster I	-	*Dickeya, Pectobacterium*
4	36H	PAS, Cache2	Periplasm, cytoplasm	0, 1, 2	Yes[c]	Class I(a) Cluster I, Class I(b), Class II, Class IV(a)	*Morganella, Yersinia*	*Dickeya, Lonsdalea, Pectobacterium*
5	36H	4HB, Cache1, Cache2	Periplasm, cytoplasm	0, 2	Yes[c]	Class I(a) Cluster I and Cluster II, Class IV(a)	*Yersinia*	*Dickeya, Pectobacterium*
6	36H	4HB	Periplasm	1, 2	Yes[c]	Class I(a) Cluster I, Class I(b)	*Hafnia, Yersinia*	*Brenneria, Dickeya, Erwinia, Enterobacter, Lonsdalea, Pectobacterium*
7	36H	4HB	Periplasm	2	Yes[c]	Class I(a) Cluster I	*Morganella*	-
8	36H	PAS	Periplasm, cytoplasm	1, 2	Yes[c]	Class I(b), Class II	*Escherichia, Hafnia, Salmonella*	*Dickeya, Erwinia, Enterobacter*

9	36H	Cache1	Periplasm	2	Yes	Class I(a) Cluster II	*Hafnia, Yersinia*	*Brenneria, Dickeya, Enterobacter, Lonsdalea, Pectobacterium*
10	36H	4HB	Periplasm	2	Yes[c]	Class I(a) Cluster I	-	*Brenneria, Dickeya, Lonsdalea, Pectobacterium*
11	24H, 28H, 36H, 44H, 52H	4HB, FIST, PAS3-PAS4, No LBD	Periplasm, cytoplasm	0, 2	Yes[c], No	Class I(a) Cluster I, Class IV(a), Class IV(b)	*Hafnia, Morganella, Salmonella, Yersinia*	*Dickeya, Lonsdalea*
12	36H	4HB, unknown	Periplasm, cytoplasm	2, 3	Yes	Class I(a)[d], Class I(a) Cluster I, Class III(c)	*Morganella, Salmonella, Yersinia*	*Brenneria, Dickeya, Enterobacter, Pectobacterium*
13	36H, unknown	PAS, No LBD	Cytoplasm	0, 2	Yes[c], No	Class II, Class IV(b)	-	*Brenneria, Dickeya, Pectobacterium*
14	36H	4HB, Cache3-2 fusion, NIT, unknown	Periplasm, cytoplasm	1, 2, 6	Yes[c]	Class I(a) Cluster I and II, Class I(b), Class III(c)	*Hafnia*	*Dickeya, Erwinia, Enterobacter, Pectobacterium*
15	36H	4HB	Periplasm, cytoplasm	2, 3	Yes	Class I(a) Cluster I, Class III(c)	*Morganella, Salmonella*	*Dickeya, Enterobacter, Lonsdalea, Pectobacterium*
16	36H	4HB, HBM	Periplasm	2	Yes	Class I(a) Cluster I and II	*Hafnia*	*Dickeya, Pectobacterium*

17	36H	4HB, unknown	Periplasm	1, 2	Yes[c]	Class I(a)[d], Class I(a) Cluster I, Class I(b)	*Morganella, Salmonella, Yersinia*	*Brenneria, Dickeya, Erwinia, Enterobacter, Lonsdalea, Pectobacterium*
18	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I	*Escherichia, Salmonella*	*Dickeya, Enterobacter, Pectobacterium*
19	36H	4HB, HBM	Periplasm	1, 2	Yes	Class I(a) Cluster I and II, Class I(b)	*Yersinia*	*Brenneria, Dickeya, Lonsdalea, Pectobacterium*
20	36H	4HB	Periplasm	2	Yes	Class I(a) Cluster I	*Morganella*	*Pectobacterium*
21	36H	Cache1	Periplasm	1, 2	Yes[c]	Class I(a) Cluster II, Class I(b)	-	*Dickeya, Pectobacterium*
22	36H, unknown	4HB	Periplasm	1, 2	Yes[c]	Class I(a) Cluster I, Class I(b)	*Escherichia, Hafnia, Morganella, Sodalis, Salmonella, Yersinia*	*Brenneria, Dickeya, Erwinia, Enterobacter, Lonsdalea, Pectobacterium*

[a]Prokaryotic MCP have diverse ligand binding domains (LBDs) such as four-helix bundle (4HB), helical bimodular (HBM), period clock protein, aryl hydrocarbon receptor and single-minded proteins (PAS), Ca^{2+} channels and chemotaxis receptors (Cache), nitrate- or nitrite- sensing domain (NIT), and F-box and intracellular signal transduction proteins (FIST) that sense broad and specific signals. 4HB domain is an ubiquitous sensory module for signal transduction in bacteria and is found to be of prokaryotic origin (24). HBM domain is made up of two four-helical bundles, of which the sequence is not identical to the sequence of 4HB (129). PAS domain is unique compared to other LBDs because it is localized in the cytosol and functions as an intracellular sensor of energy level change in a cell (130). The Cache domain is also found in eukaryotic Ca^{2+} channels in addition to prokaryotic MCP, and is thought to bind to small molecules (131).

[b]Initially, MCP were categorized into four major classes (I-IV) based on their ligand binding domain and membrane topology (30). Class III is subdivided into III(m) and III(c) based on the presence of the ligand binding domain in the later subclass (12). Lacal et al.

(29) further subdivided class I based on the number of transmembrane helices (subclasses I(a) and I(b)), and class IV based on the existence of the LBD (subclasses IV(a) and IV(b)). Subclasses I(a) and I(b) have two transmembrane helices and one transmembrane helix, respectively (29). Subclass IV(a) has a ligand binding domain, while subclass IV(b) lacks a ligand binding domain (29). In addition, class I(a) was separated into two clusters (clusters I and II) based on the sequence length of the ligand binding domain (29). Cluster I has ligand binding domain with approximately 120 - 215 amino acids (aa), while cluster II has ligand binding domain of about 215 - 299 aa (29).

[c]HAMP presence was identified by manual inspection of unidentified short region between the ligand binding domain (LBD) and the cytoplasmic signaling domain (SD) sequences.

[d]Subcluster could not be determined due to unknown ligand binding domain.

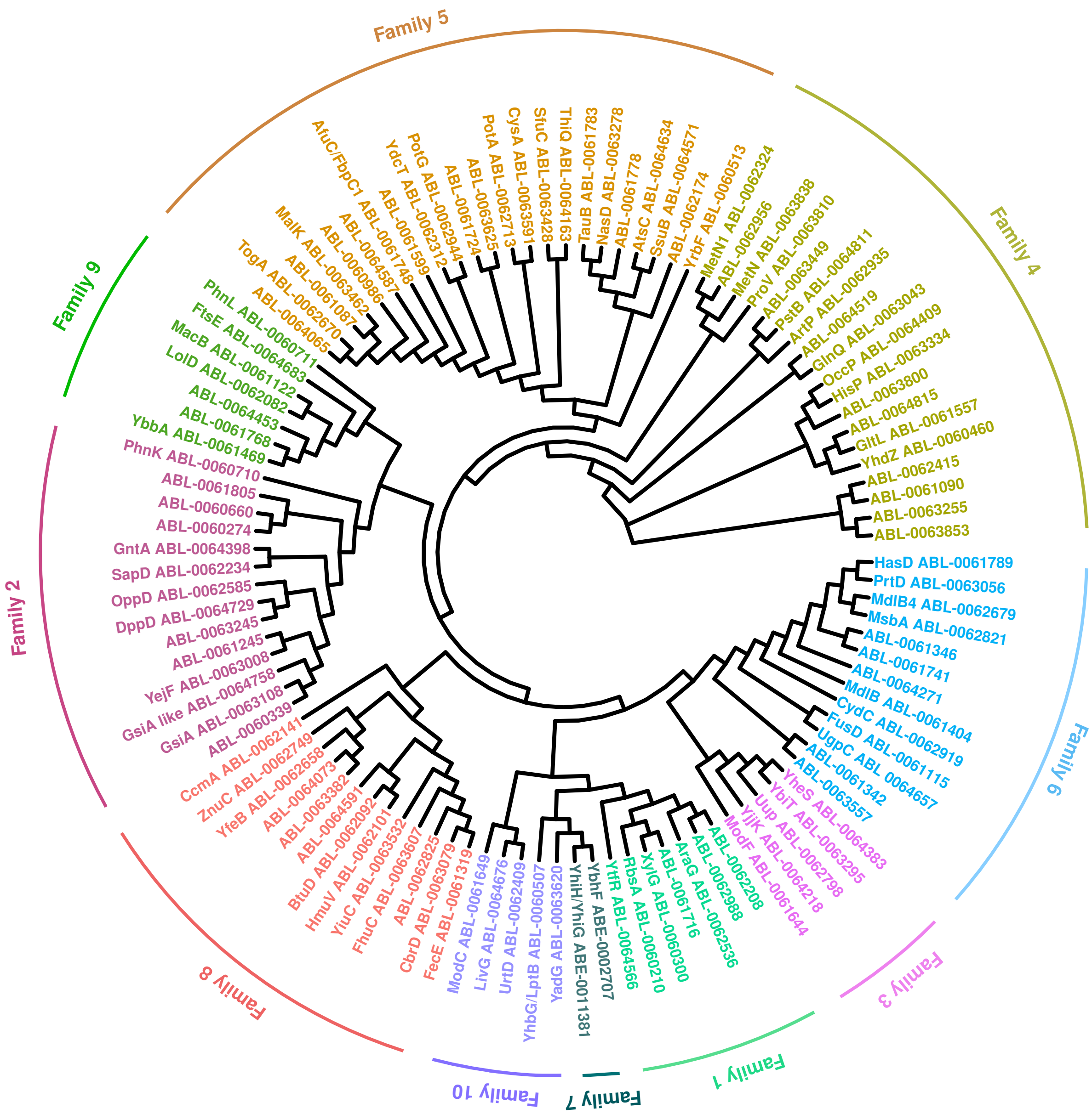

Figure 2.1: Phylogenetic tree of the NBDs representing 108 ABC transporters in *P. atrosepticum* SCRI 1043 and *Escherichia coli* K-12 MG1655**.** The evolutionary history was using the maximum likelihood method based on the JTT matrix-based model with bootstrap inference with 500 replicates. The analysis involved 108 amino acid sequences of NBDs.

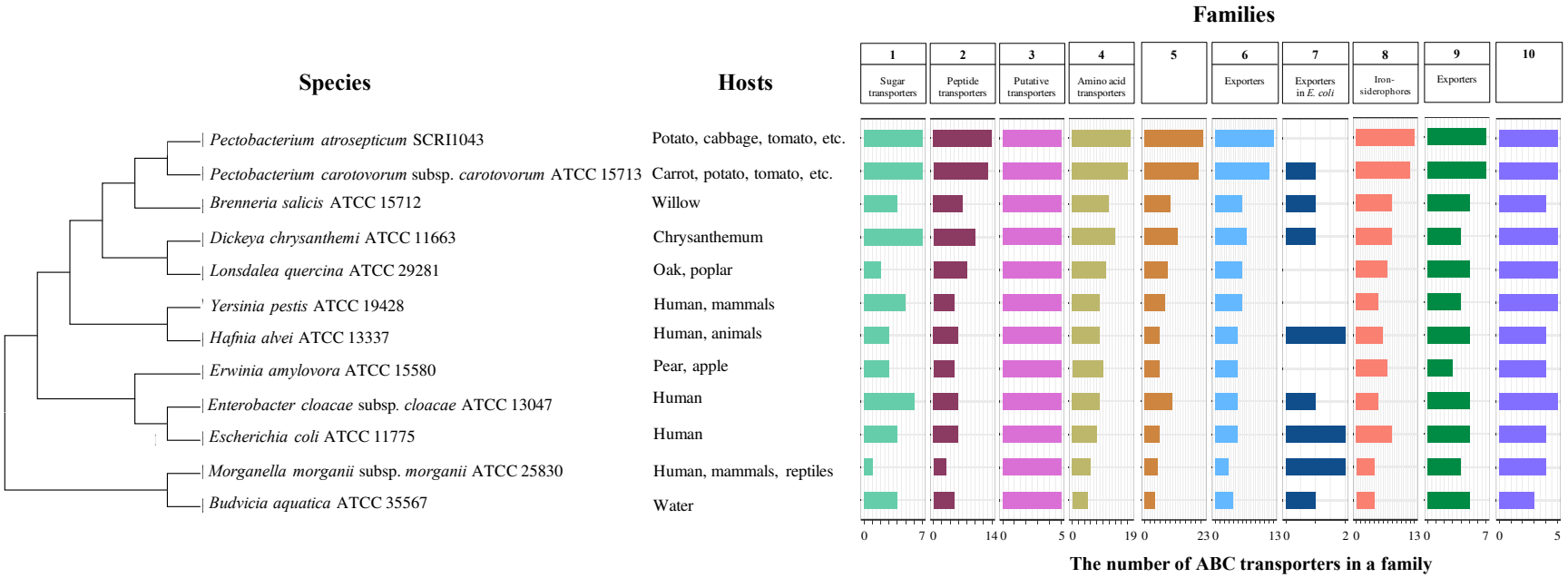

Figure 2.2: Summary of ATP-binding proteins in the order of Enterobacteriales. The tree shows phylogenetic relations of various bacteria in the order, based on the housekeeping gene *dnaX* sequence in each bacterium. Bar graphs show the number of ABC transporters that identified from the bacterial genome across different transporter families, and the color used to differentiate ABC transporter family corresponds to what is used in Figure 2.1.

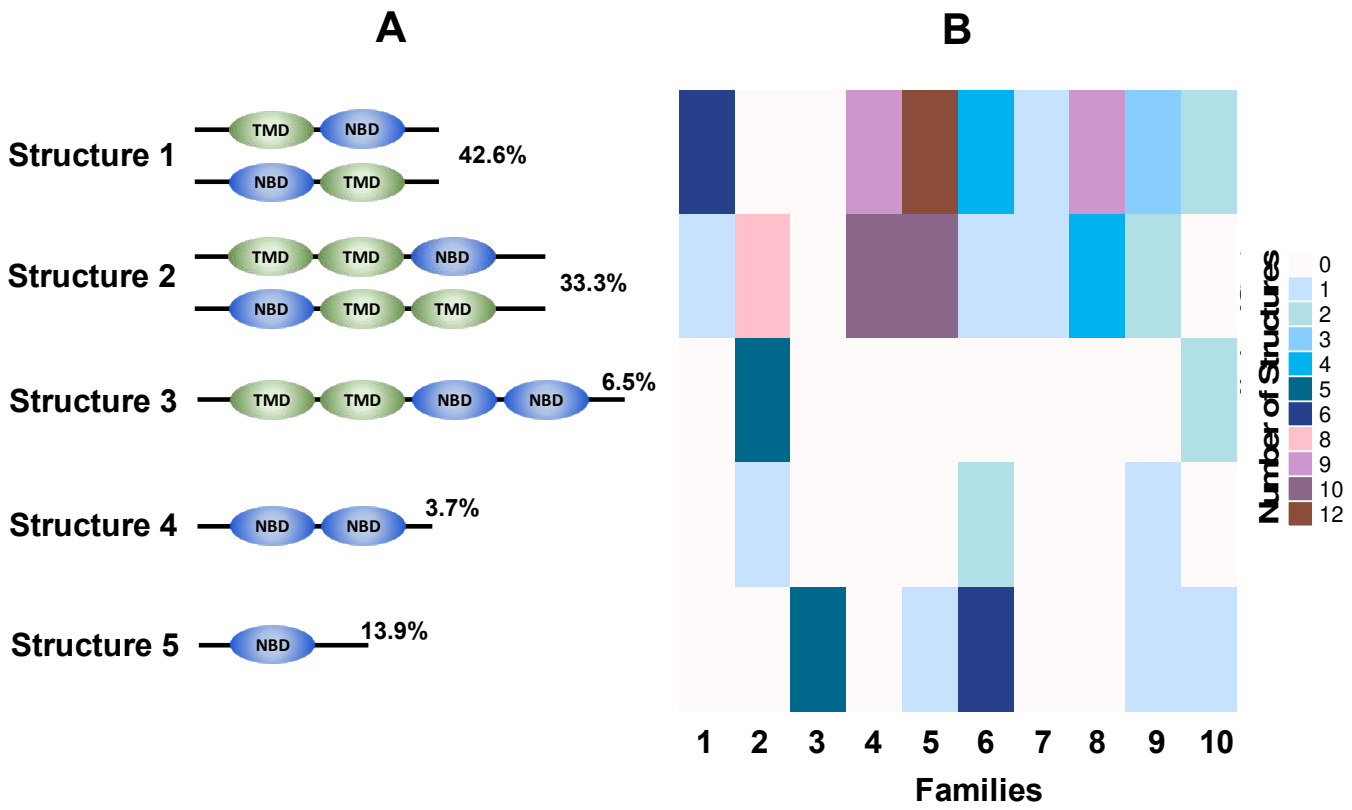

Figure 2.3: (A) Summary of common structures found in the Enterobacteriales ABC transporters, with the percentage of each structure being listed. (B) a heatmap showing the number of different structures grouped by each ABC transporter family. Each row in the heatmap corresponds to the specific structure illustrated in (A), and each column represents an ABC transporter family listed in Table 2.2 and Figure 2.1.

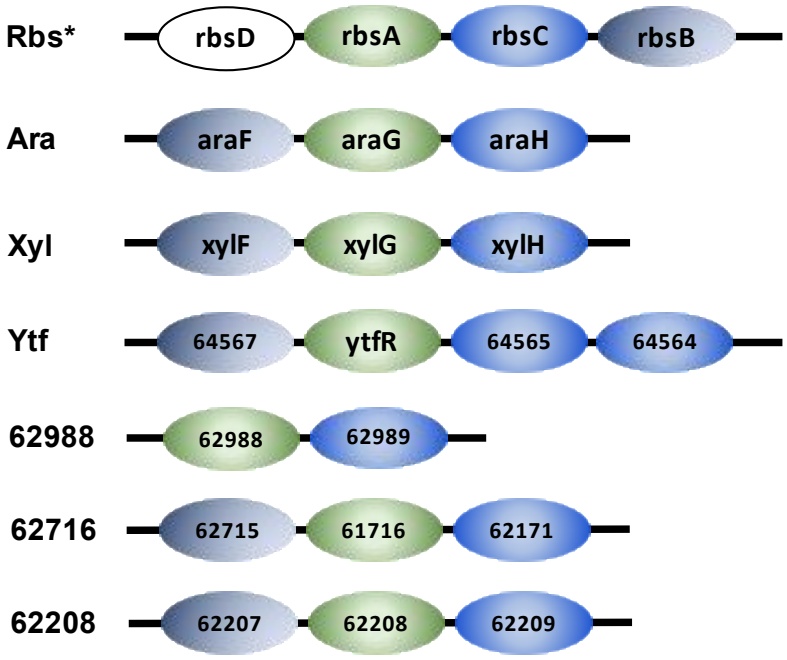

Figure 2.4: Domain compositions of ABC transporters in family 1. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, and a silver gray is a substrate binding domain. * indicate a conserved ABC transporter found in every strain.

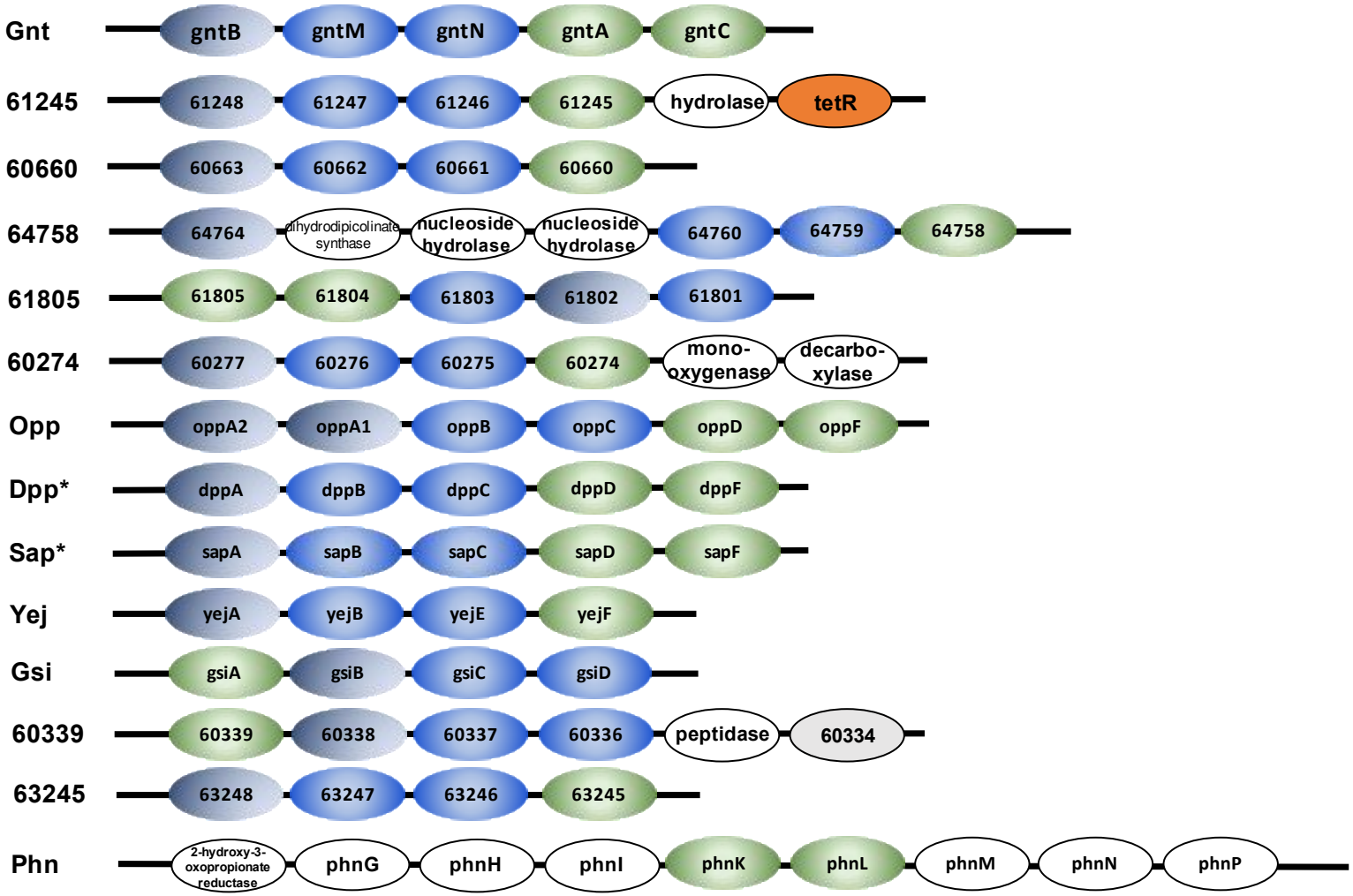

Figure 2.5: Domain compositions of ABC transporters in family 2. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, a silver gray is a substrate binding domain, a gray is a hypothetical protein, and an orange is a transcriptional regulator. * indicate a conserved ABC transporter found in every strain.

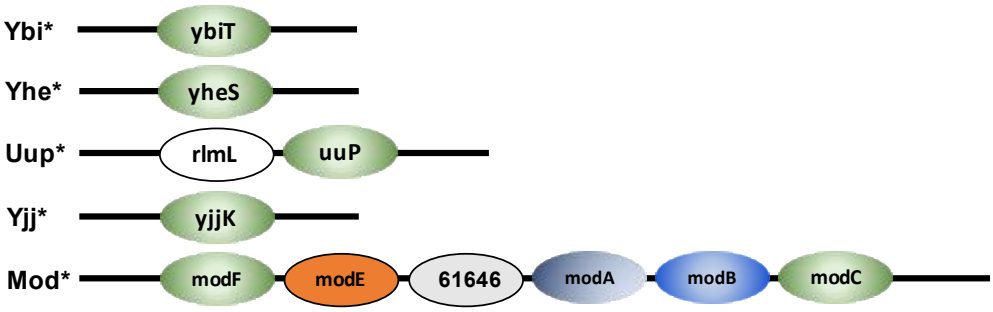

Figure 2.6: Domain compositions of ABC transporters in family 3. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, a silver gray is a substrate binding domain, a gray is a hypothetical protein, and an orange is a transcriptional regulator. * indicate a conserved ABC transporter found in every strain.

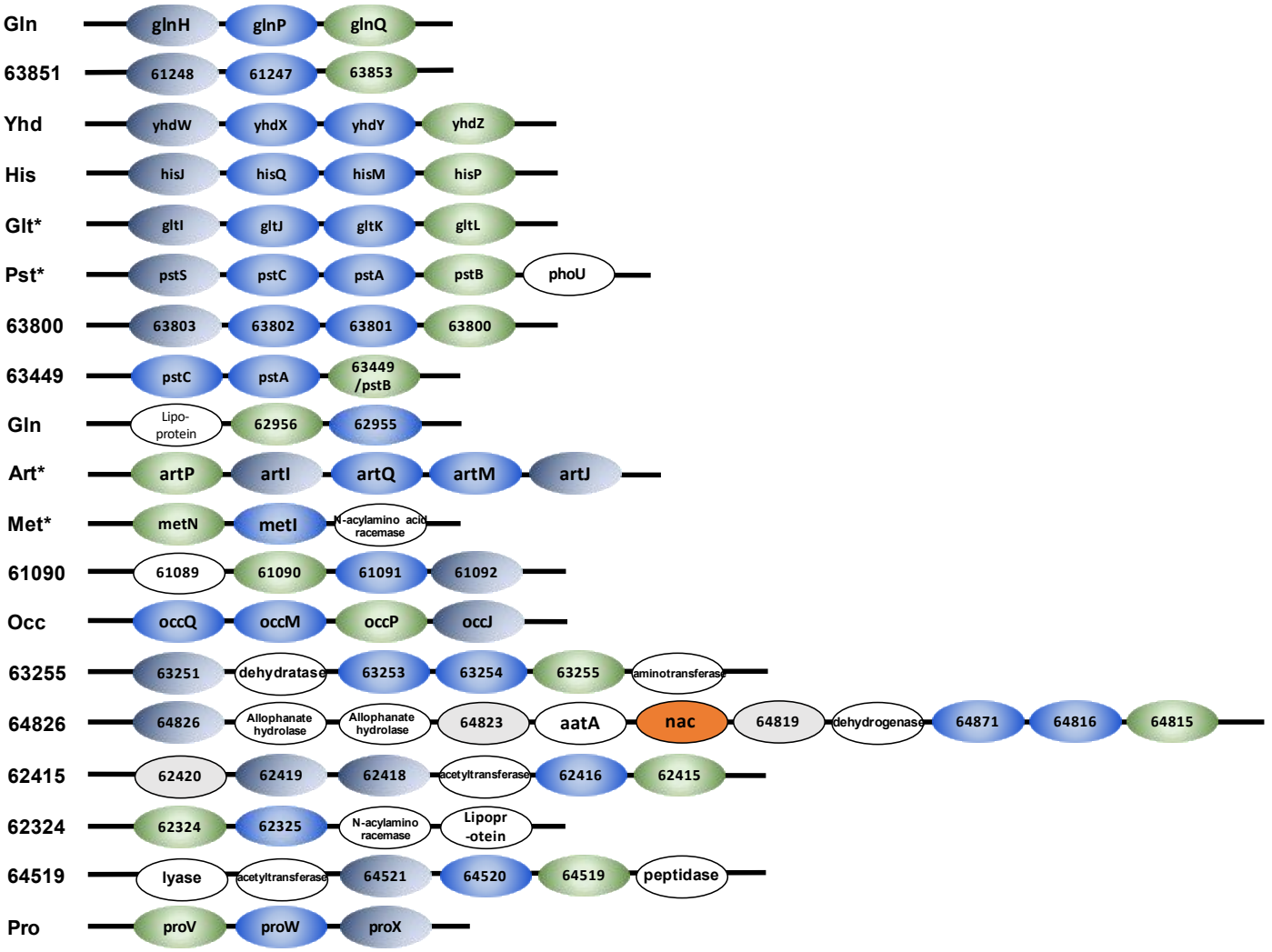

Figure 2.7: Domain compositions of ABC transporters in family 4. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, a silver gray is a substrate binding domain, a gray is a hypothetical protein, and an orange is a transcriptional regulator. * indicate a conserved ABC transporter found in every strain.

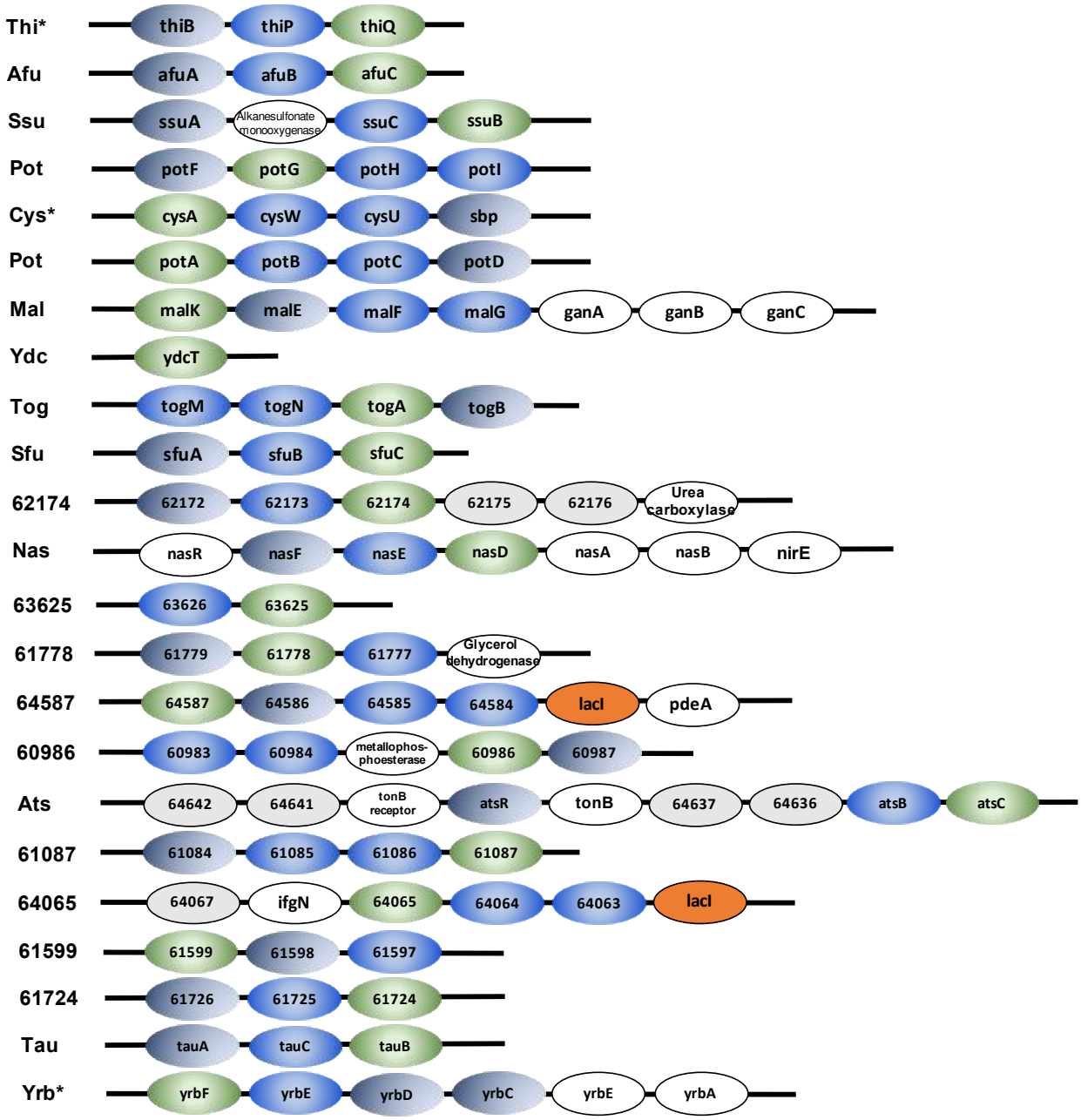

Figure 2.8: Domain compositions of ABC transporters in family 5. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, a silver gray is a substrate binding domain, a gray is a hypothetical protein, and an orange is a transcriptional regulator. * indicate a conserved ABC transporter found in every strain.

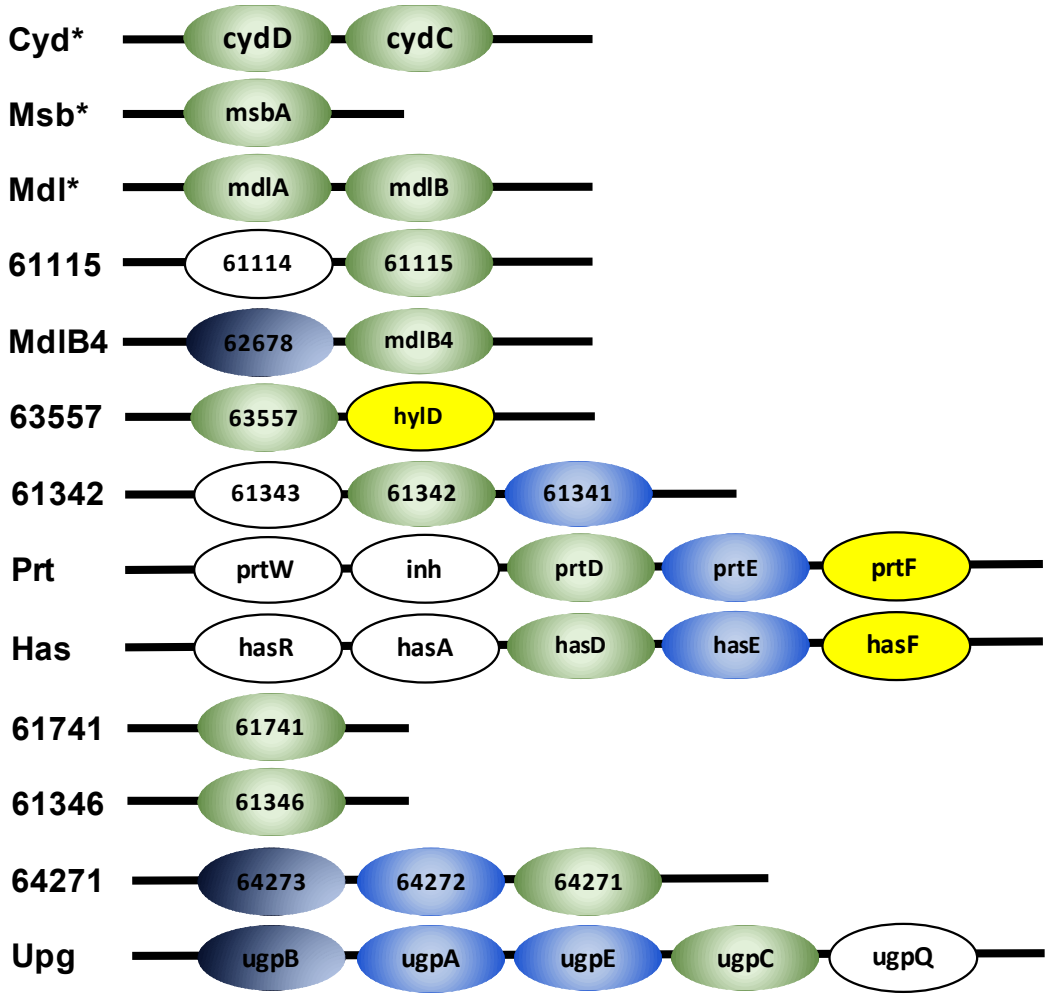

Figure 2.9: Domain compositions of ABC transporters in family 6. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, a silver gray is a substrate binding domain, and a yellow is an outer membrane. * indicate a conserved ABC transporter found in every strain.

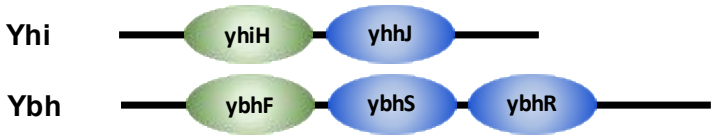

Figure 2.10: Domain compositions of ABC transporters in family 7. A green is an ATP-binding domain and a blue is a transmembrane domain.

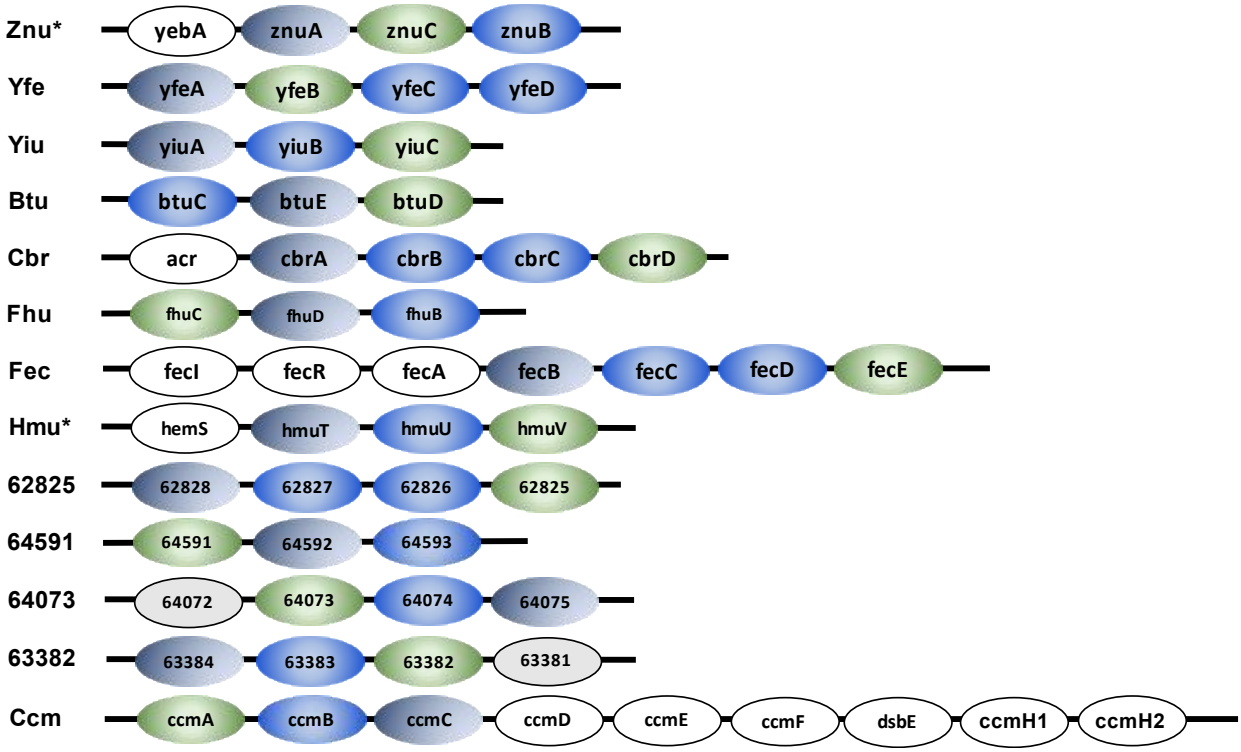

Figure 2.11: Domain compositions of ABC transporters in family 8. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, and a silver gray is a substrate binding domain. * indicate a conserved ABC transporter found in every strain.

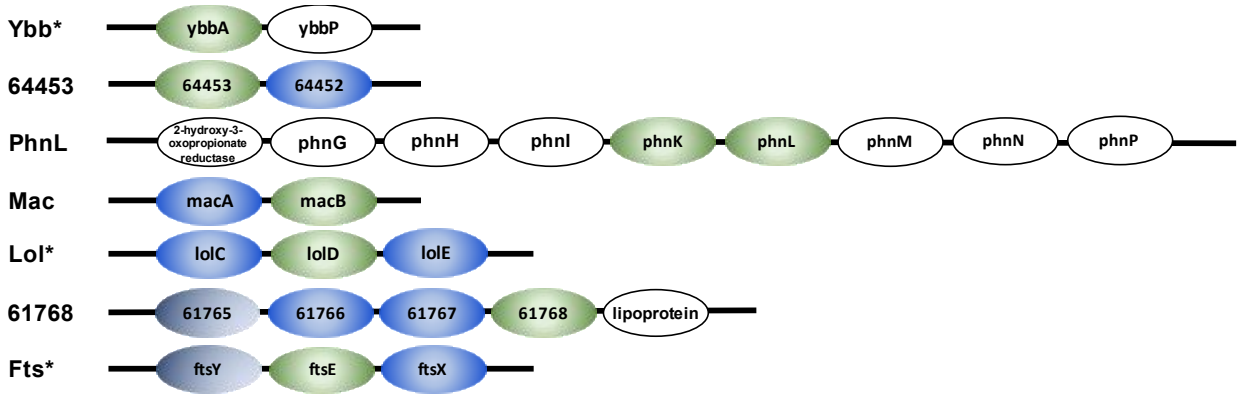

Figure 2.12: Domain compositions of ABC transporters in family 9. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, and a silver gray is a substrate binding domain. * indicate a conserved ABC transporter found in every strain.

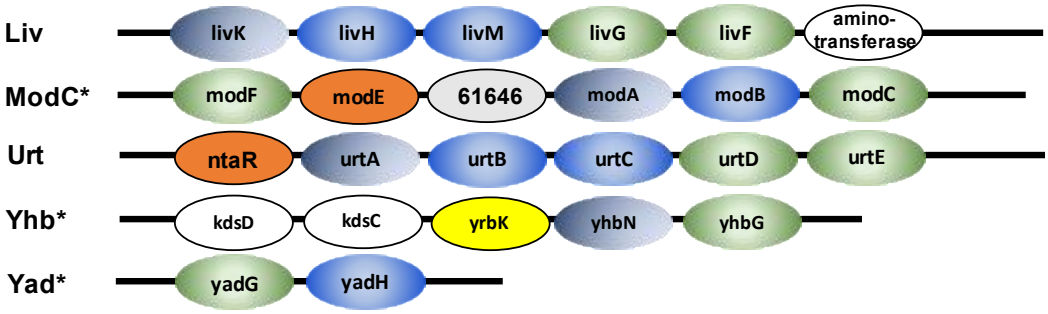

Figure 2.13: Domain compositions of ABC transporters in family 10. A white oval indicates a protein or an enzyme, a green is an ATP-binding domain, a blue is a transmembrane domain, and a silver gray is a substrate binding domain, a yellow is an outer membrane, and an orange is a transcriptional regulator. * indicate a conserved ABC transporter found in every strain.

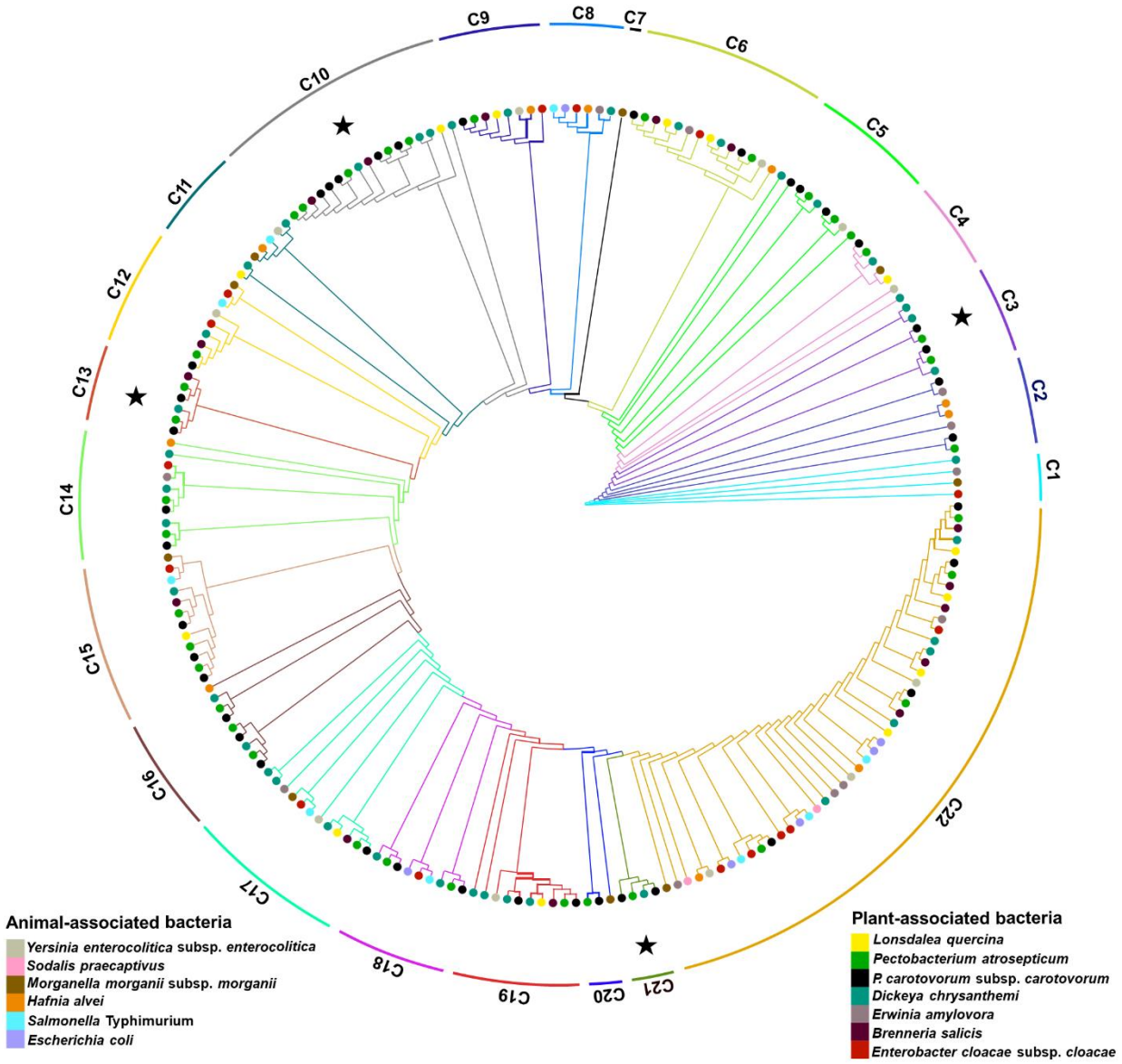

Figure 2.14: A maximum likelihood phylogenetic tree built from concatenated alignment of 217 MCP cytoplasmic signaling domain protein sequences from 13 representative strains in the order Enterobacteriales. *Budvicia* is not included because it does not contain any *mcp* in its genome. Each clade is indicated by a different color and a capital letter 'C'. A 'star' indicates clade with exclusively plant-associated bacteria. The clades were determined based on overall similarity and consistency of the corresponding ligand binding domain, number of transmembrane helix region, and the heptad nomenclature of the signaling domain. The branching patterns and lengths were also considered when determining the clades.

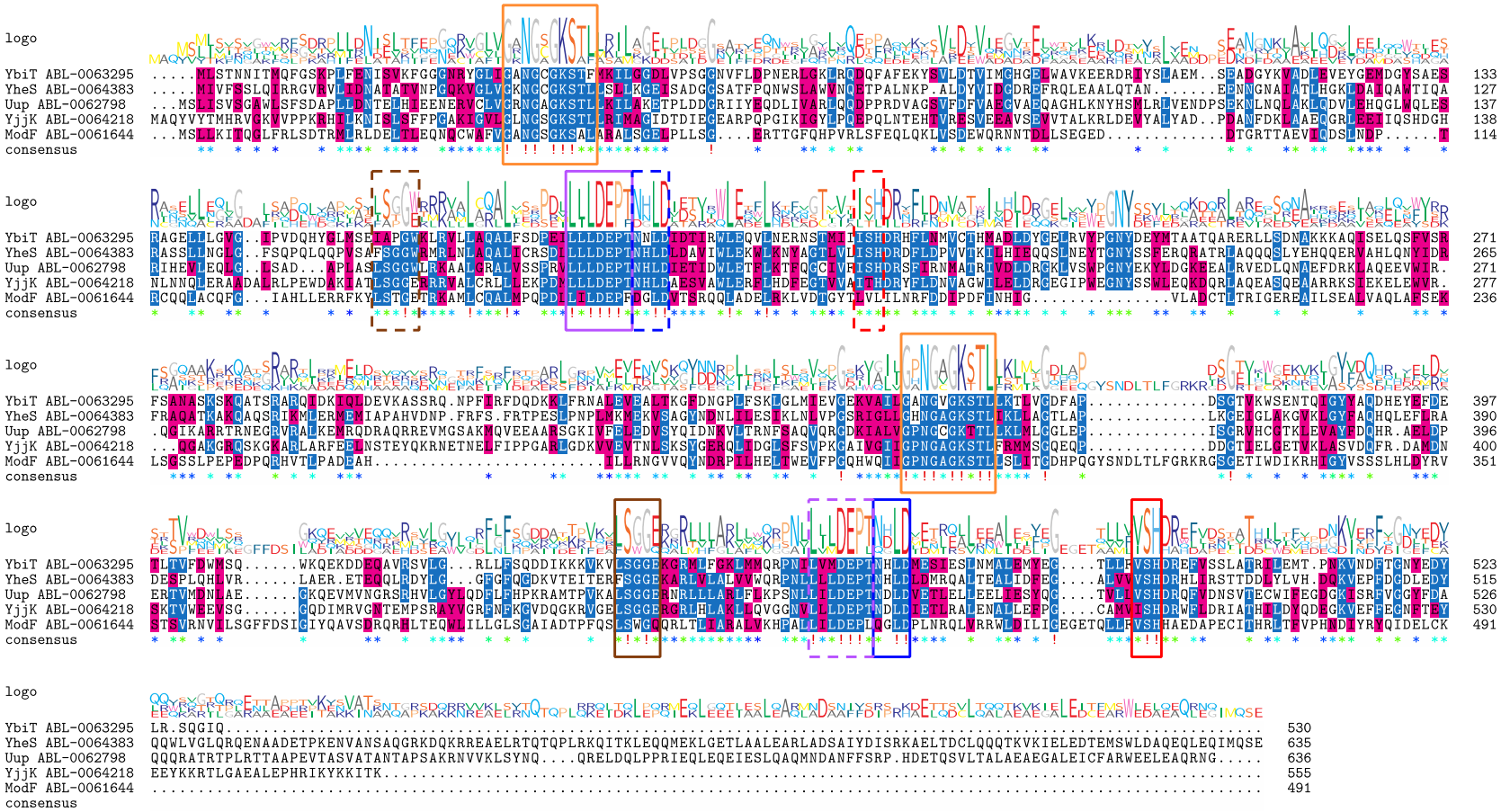

Figure 2.15: Sequence conservation of NBDs in ABC transporter family 3. The Walker A (P-loop) is orange, the Walker B purple, the H-motif red, the signature motif brown, and the D-loop blue. The putative motifs are labeled with dashed boxes.

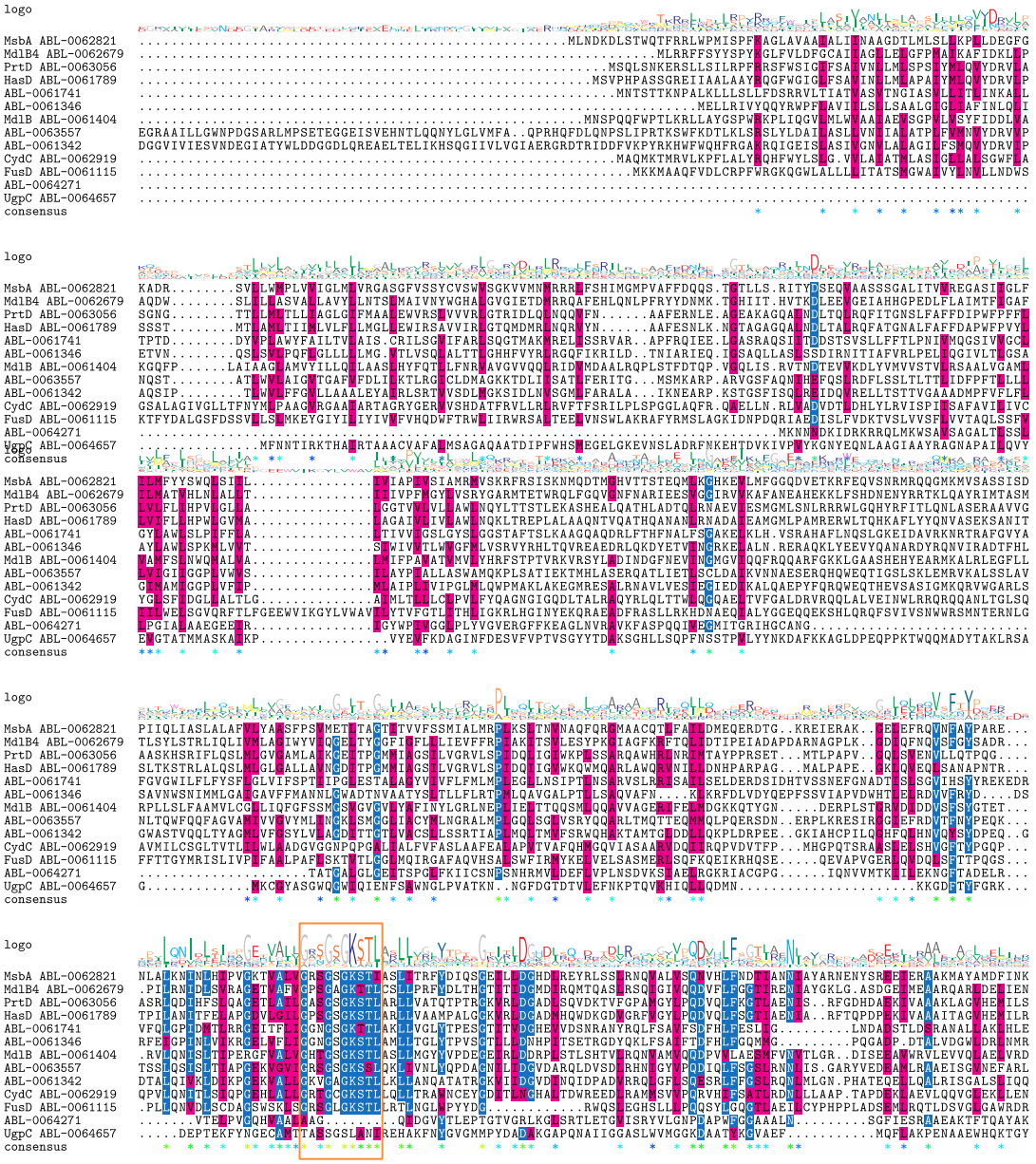

Figure 2.16: Sequence conservation of NBDs in ABC transporter family 6, only showing regions with observed conservation. The Walker A (P-loop) is orange, the Walker B purple, the H-motif red, and the signature motif brown, and the D-loop blue.

REFERENCES

1. Rees DC, Johnson E, Lewinson O. ABC transporters: the power to change. Nature Reviews Molecular Cell Biology. 2009;10(3):218-27.

2. Hugouvieux-Cotte-Pattat N, Blot N, Reverchon S. Identification of TogMNAB, an ABC transporter which mediates the uptake of pectic oligomers in *Erwinia chrysanthemi* 3937. Molecular Microbiology. 2001;41(5):1113-23.

3. Konishi H, Hio M, Kobayashi M, Takase R, Hashimoto W. Bacterial chemotaxis towards polysaccharide pectin by pectin-binding protein. Scientific Reports. 2020;10.

4. Hegde M, Englert DL, Schrock S, Cohn WB, Vogt C, Wood TK, et al. Chemotaxis to the quorum-sensing signal AI-2 requires the Tsr chemoreceptor and the periplasmic LsrB AI-2-binding protein. Journal of Bacteriology. 2011;193(3):768-73.

5. Ud-Din AIMS, Roujeinikova A. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. Cellular and Molecular Life Sciences. 2017;74(18):3293-303.

6. Wuichet K, Zhulin IB. Origins and diversification of a complex signal transduction system in prokaryotes. Science Signaling. 2010;3(128):ra50-ra.

7. Zeng Y, Charkowski AO. The role of ATP-binding cassette (ABC) transporters in bacterial phytopathogenesis. Phytopathology. 2020.

8. Linton KJ, Higgins CF. The *Escherichia coli* ATP-binding cassette (ABC) proteins. Molecular Microbiology. 1998;28(1):5-13.

9. Moussatova A, Kandt C, O'Mara ML, Tieleman DP. ATP-binding cassette transporters in *Escherichia coli*. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2008;1778(9):1757-71.

10. Wilkens S. Structure and mechanism of ABC transporters. F1000Prime Reports. 2015;7.

11. Berntsson RP-A, Smits SH, Schmitt L, Slotboom D-J, Poolman B. A structural classification of substrate-binding proteins. FEBS Letters. 2010;584(12):2606-17.

12. Wuichet K, Alexander RP, Zhulin IB. Comparative genomic and protein sequence analyses of a complex system controlling bacterial chemotaxis. Methods in Enzymology. 422: Elsevier; 2007. p. 3-31.

13. Wadhams GH, Armitage JP. Making sense of it all: bacterial chemotaxis. Nature Reviews Molecular Cell Biology. 2004;5(12):1024.

14. Adeolu M, Alnajar S, Naushad S, Gupta RS. Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae, Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. International Journal of Systematic and Evolutionary Microbiology. 2016;66(12):5575-99.

15. Glasner JD, Liss P, Plunkett III G, Darling A, Prasad T, Rusch M, et al. ASAP, a systematic annotation package for community analysis of genomes. Nucleic Acids Research. 2003;31(1):147-51.

16. Glasner JD, Rusch M, Liss P, Plunkett III G, Cabot EL, Darling A, et al. ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Research. 2006;34(suppl_1):D41-D5.

17. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8.

18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution. 2018;35(6):1547-9.

19. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, et al. MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Research. 2007;35(suppl_2):W433-W7.

20. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Research. 2021;49(D1):D344-D54.

21. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Research. 2018;46(D1):D493-D6.

22. Gumerov VM, Ortega DR, Adebali O, Ulrich LE, Zhulin IB. MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. Nucleic Acids Research. 2020;48(D1):D459-D64.

23. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Research. 2019;47(D1):D427-D32.

24. Ulrich LE, Zhulin IB. Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. Bioinformatics. 2005;21(Suppl_3):iii45-iii8.

25. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology. 2001;305(3):567-80.

26. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 2011;8(10):785-6.

27. Cserzo M, Eisenhaber F, Eisenhaber B, Simon I. TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. Bioinformatics. 2004;20(1):136-7.

28. Cserzö M, Eisenhaber F, Eisenhaber B, Simon I. On filtering false positive transmembrane protein predictions. Protein Engineering, Design and Selection. 2002;15(9):745-52.

29. Lacal J, García-Fontana C, Muñoz-Martínez F, Ramos JL, Krell T. Sensing of environmental signals: classification of chemoreceptors according to the size of their ligand binding regions. Environmental Microbiology. 2010;12(11):2873-84.

30. Zhulin IB. The superfamily of chemotaxis transducers: from physiology to genomics and back. Advances in Microbial Physiology. 2001;45:157-98.

31. Gumerov VM, Ortega DR, Adebali O, Ulrich LE, Zhulin IB. MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. Nucleic Acids Research. 2019;48(D1):D459-D64.

32. Alexander RP, Zhulin IB. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(8):2885-90.

33. Dale C, Young SA, Haydon DT, Welburn SC. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(4):1883-8.

34. Charkowski A, Blanco C, Condemine G, Expert D, Franza T, Hayes C, et al. The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. Annual Review of Phytopathology. 2012;50:425-49.

35. Ma B, Hibbing ME, Kim H-S, Reedy RM, Yedidia I, Breuer J, et al. Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. Phytopathology. 2007;97(9):1150-63.

36. Holtsmark I, Eijsink VG, Brurberg MB. Bacteriocins from plant pathogenic bacteria. FEMS Microbiology Letters. 2008;280(1):1-7.

37. Lewis VG, Ween MP, McDevitt CA. The role of ATP-binding cassette transporters in bacterial pathogenicity. Protoplasma. 2012;249(4):919-42.

38. Tanaka KJ, Song S, Mason K, Pinkett HW. Selective substrate uptake: the role of ATP-binding cassette (ABC) importers in pathogenesis. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2018;1860(4):868-77.

39. Matilla MA, Krell T. The effect of bacterial chemotaxis on host infection and pathogenicity. FEMS Microbiology Reviews. 2017;42(1):fux052.

40. Río-Álvarez I, Muñoz-Gómez C, Navas-Vásquez M, Martínez-García PM, Antúnez-Lamas M, Rodríguez-Palenzuela P, et al. Role of *Dickeya dadantii* 3937 chemoreceptors in the entry to Arabidopsis leaves through wounds. Molecular Plant Pathology. 2015;16(7):685-98.

41. Duan X, Quiocho FA. Structural evidence for a dominant role of nonpolar interactions in the binding of a transport/chemosensory receptor to its highly polar ligands. Biochemistry. 2002;41(3):706-12.

42. Becq J, Churlaud C, Deschavanne P. A benchmark of parametric methods for horizontal transfers detection. PLOS ONE. 2010;5(4):e9989.

43. Hamady M, Betterton M, Knight R. Using the nucleotide substitution rate matrix to detect horizontal gene transfer. BMC Bioinformatics. 2006;7(1):476.

44. Xiong J, Feng J, Yuan D, Zhou J, Miao W. Tracing the structural evolution of eukaryotic ATP binding cassette transporter superfamily. Scientific Reports. 2015;5:16724.

45. Lewinson O, Livnat-Levanon N. Mechanism of action of ABC importers: conservation, divergence, and physiological adaptations. Journal of Molecular Biology. 2017;429(5):606-19.

46. Morgan JL, Acheson JF, Zimmer J. Structure of a type-1 secretion system ABC transporter. Structure. 2017;25(3):522-9.

47. Chekabab SM, Harel J, Dozois CM. Interplay between genetic regulation of phosphate homeostasis and bacterial virulence. Virulence. 2014;5(8):786-93.

48. Franza T, Expert D. Role of iron homeostasis in the virulence of phytopathogenic bacteria: an ‘a la carte’menu. Molecular Plant Pathology. 2013;14(4):429-38.

49. McCraw S, Park D, Jones R, Bentley M, Rico A, Ratcliffe R, et al. GABA (γ-aminobutyric acid) uptake via the GABA permease GabP represses virulence gene expression in *Pseudomonas syringae* pv. *tomato* DC3000. Molecular Plant-Microbe Interactions. 2016;29(12):938-49.

50. Sharma A, Sharma D, Verma SK. Zinc binding proteome of a phytopathogen *Xanthomonas translucens* pv. *undulosa*. Royal Society Open Science. 2019;6(9):190369.

51. Fatima U, Senthil-Kumar M. Plant and pathogen nutrient acquisition strategies. Frontiers in Plant Science. 2015;6:750.

52. Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, et al. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science. 2001;294(5550):2317-23.

53. Bouvet OM, Grimont PA, Richard C, Aldova E, Hausner O, Gabrhelova M. *Budvicia aquatica* gen. nov., sp. nov.: a hydrogen sulfide-producing member of the *Enterobacteriaceae*. International Journal of Systematic and Evolutionary Microbiology. 1985;35(1):60-4.

54. Ortega DR, Zhulin IB. Evolutionary genomics suggests that CheV is an additional adaptor for accommodating specific chemoreceptors within the chemotaxis signaling complex. PLOS Computational Biology. 2016;12(2):e1004723.

55. Ma B, Charkowski AO, Glasner JD, Perna NT. Identification of host-microbe interaction factors in the genomes of soft rot-associated pathogens *Dickeya dadantii* 3937 and *Pectobacterium carotovorum* WPP14 with supervised machine learning. BMC Genomics. 2014;15(1):508.

56. Brenner DJ, McWhorter AC, Kai A, Steigerwalt AG, Farmer J. *Enterobacter asburiae* sp. nov., a new species found in clinical specimens, and reassignment of *Erwinia dissolvens* and *Erwinia nimipressuralis* to the genus *Enterobacter* as *Enterobacter dissolvens* comb. nov. and *Enterobacter nimipressuralis* comb. nov. Journal of Clinical Microbiology. 1986;23(6):1114-20.

57. Doijad S, Imirzalioglu C, Yao Y, Pati NB, Falgenhauer L, Hain T, et al. *Enterobacter bugandensis* sp. nov., isolated from neonatal blood. International Journal of Systematic and Evolutionary Microbiology. 2016;66(2):968-74.

58. Ren Y, Ren Y, Zhou Z, Guo X, Li Y, Feng L, et al. Complete genome sequence of *Enterobacter cloacae* subsp. *cloacae* type strain ATCC 13047. Journal of Bacteriology. 2010;192(9):2463-4.

59. Hoffmann H, Stindl S, Ludwig W, Stumpf A, Mehlen A, Monget D, et al. *Enterobacter hormaechei* subsp. *oharae* subsp. nov., *E. hormaechei* subsp. *hormaechei* comb. nov., and *E.*

hormaechei subsp. *steigerwaltii* subsp. nov., three new subspecies of clinical importance. Journal of Clinical Microbiology. 2005;43(7):3297-303.

60. Hoffmann H, Stindl S, Stumpf A, Mehlen A, Monget D, Heesemann J, et al. Description of *Enterobacter ludwigii* sp. nov., a novel *Enterobacter* species of clinical relevance. Systematic and Applied Microbiology. 2005;28(3):206-12.

61. Zhu B, Zhang G-Q, Lou M-M, Tian W-X, Li B, Zhou X-P, et al. Genome sequence of the *Enterobacter mori* type strain, LMG 25706, a pathogenic bacterium of *Morus alba* L. Journal of Bacteriology. 2011.

62. Hoffmann H, Roggenkamp A. Population genetics of the nomenspecies *Enterobacter cloacae*. Applied and Environmental Microbiology. 2003;69(9):5306-18.

63. Sutton GG, Brinkac LM, Clarke TH, Fouts DE. *Enterobacter hormaechei* subsp. *hoffmannii* subsp. nov., *Enterobacter hormaechei* subsp. *xiangfangensis* comb. nov., *Enterobacter roggenkampii* sp. nov., and *Enterobacter muelleri* is a later heterotypic synonym of *Enterobacter asburiae* based on computational analysis of sequenced *Enterobacter* genomes. F1000Research. 2018;7.

64. Manter DK, Hunter WJ, Vivanco JM. *Enterobacter soli* sp. nov.: a lignin-degrading γ-proteobacteria isolated from soil. Current Microbiology. 2011;62(3):1044-9.

65. Huys G, Cnockaert M, Janda JM, Swings J. *Escherichia albertii* sp. nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. International Journal of Systematic and Evolutionary Microbiology. 2003;53(3):807-10.

66. Wadley TD, Jenjaroenpun P, Wongsurawat T, Ussery DW, Nookaew I. Complete genome and plasmid sequences of *Escherichia coli* type strain ATCC 11775. Microbiology Resource Announcements. 2019;8(9):e00046-19.

67. Farmer J, Fanning G, Davis B, O'hara C, Riddle C, Hickman-Brenner F, et al. *Escherichia fergusonii* and *Enterobacter taylorae*, two new species of *Enterobacteriaceae* isolated from clinical specimens. Journal of Clinical Microbiology. 1985;21(1):77-81.

68. Liu S, Jin D, Lan R, Wang Y, Meng Q, Dai H, et al. *Escherichia marmotae* sp. nov., isolated from faeces of *Marmota himalayana*. International Journal of Systematic and Evolutionary Microbiology. 2015;65(7):2130-4.

69. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, et al. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature. 2001;413(6858):852-6.

70. Mizuno A, Shigeyoshi S, Kawai A, Nishiyama K. Taxonomic position of the causal pathogen of bacterial shoot blight of pear. Journal of General Plant Pathology. 2000;66(1):48-58.

71. Mergaert J, Hauben L, Cnockaert MC, Swings J. Reclassification of non-pigmented *Erwinia herbicola* strains from trees as *Erwinia billingiae* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 1999;49(2):377-83.

72. Kube M, Migdoll AM, Gehring I, Heitmann K, Mayer Y, Kuhl H, et al. Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*. BMC Genomics. 2010;11(1):393.

73. Rezzonico F, Smits TH, Born Y, Blom J, Frey JE, Goesmann A, et al. *Erwinia gerundensis* sp. nov., a cosmopolitan epiphyte originally isolated from pome fruit trees. International Journal of Systematic and Evolutionary Microbiology. 2016;66(3):1583-92.

74. Campillo T, Luna E, Portier P, Fischer-Le Saux M, Lapitan N, Tisserat NA, et al. *Erwinia iniecta* sp. nov., isolated from Russian wheat aphid (*Diuraphis noxia*). International Journal of Systematic and Evolutionary Microbiology. 2015;65(10):3625-33.

75. Moretti C, Hosni T, Vandemeulebroecke K, Brady C, De Vos P, Buonaurio R, et al. *Erwinia oleae* sp. nov., isolated from olive knots caused by *Pseudomonas savastanoi* pv. *savastanoi*. International Journal of Systematic and Evolutionary Microbiology. 2011;61(11):2745-52.

76. Hao M, BRENNER DJ, STEIGERWALT AG, Kosako Y, Komagata K. *Erwinia persicinus*, a new species isolated from plants. International Journal of Systematic and Evolutionary Microbiology. 1990;40(4):379-83.

77. Mohamaden WI, Zhen-fen Z, Hegab IM, Shang-li S. Experimental infection in mice with *Erwinia persicina*. Microbial Pathogenesis. 2019;130:38-43.

78. López MM, Rosello M, Llop P, Ferrer S, Christen R, Gardan L. *Erwinia piriflorinigrans* sp. nov., a novel pathogen that causes necrosis of pear blossoms. International Journal of Systematic and Evolutionary Microbiology. 2011;61(3):561-7.

79. Smits TH, Rezzonico F, López MM, Blom J, Goesmann A, Frey JE, et al. Phylogenetic position and virulence apparatus of the pear flower necrosis pathogen *Erwinia piriflorinigrans* CFBP 5888T as assessed by comparative genomics. Systematic and Applied Microbiology. 2013;36(7):449-56.

80. Kim W-S, Gardan L, Rhim S-L, Geider K. *Erwinia pyrifoliae* sp. nov., a novel pathogen that affects Asian pear trees (*Pyrus pyrifolia* Nakai). International Journal of Systematic and Evolutionary Microbiology. 1999;49(2):899-906.

81. Kube M, Migdoll AM, Müller I, Kuhl H, Beck A, Reinhardt R, et al. The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. Environmental Microbiology. 2008;10(9):2211-22.

82. Liu B, Luo J, Li W, Long X-F, Zhang Y-Q, Zeng Z-G, et al. *Erwinia teleogrylli* sp. nov., a bacterial isolate associated with a Chinese cricket. PLOS ONE. 2016;11(1):e0146596.

83. Bell K, Sebaihia M, Pritchard L, Holden M, Hyman L, Holeva M, et al. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(30):11105-10.

84. Gardan L, Gouy C, Christen R, Samson R. Elevation of three subspecies of *Pectobacterium carotovorum* to species level: *Pectobacterium atrosepticum* sp. nov., *Pectobacterium betavasculorum* sp. nov. and *Pectobacterium wasabiae* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2003;53(2):381-91.

85. Duarte V, De Boer S, Ward L, De Oliveira A. Characterization of atypical *Erwinia carotovora* strains causing blackleg of potato in Brazil. Journal of Applied Microbiology. 2004;96(3):535-45.

86. Waleron M, Waleron K, Lojkowska E. Characterization of *Pectobacterium carotovorum* subsp. *odoriferum* causing soft rot of stored vegetables. European Journal of Plant Pathology. 2014;139(3):457-69.

87. Khayi S, Cigna J, Chong TM, Quêtu-Laurent A, Chan K-G, Hélias V, et al. Transfer of the potato plant isolates of *Pectobacterium wasabiae* to *Pectobacterium parmentieri* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2016;66(12):5379-83.

88. Zoledowska S, Motyka-Pomagruk A, Sledz W, Mengoni A, Lojkowska E. High genomic variability in the plant pathogenic bacterium *Pectobacterium parmentieri* deciphered from de novo assembled complete genomes. BMC Genomics. 2018;19(1):751.

89. Dees MW, Lysøe E, Rossmann S, Perminow J, Brurberg MB. *Pectobacterium polaris* sp. nov., isolated from potato (*Solanum tuberosum*). International Journal of Systematic and Evolutionary Microbiology. 2017;67(12):5222-9.

90. Goto M, Matsumoto K. *Erwinia carotovora* subsp. *wasabiae* subsp. nov. isolated from diseased rhizomes and fibrous roots of Japanese horseradish (*Eutrema wasabi* Maxim.). International Journal of Systematic and Evolutionary Microbiology. 1987;37(2):130-5.

91. Parkinson N, DeVos P, Pirhonen M, Elphinstone J. *Dickeya aquatica* sp. nov., isolated from waterways. International Journal of Systematic and Evolutionary Microbiology. 2014;64(7):2264-6.

92. Pritchard L, Humphris S, Saddler GS, Elphinstone JG, Pirhonen M, Toth IK. Draft genome sequences of 17 isolates of the plant pathogenic bacterium *Dickeya*. Genome Announcements. 2013;1(6):e00978-13.

93. Bull CT, Coutinho T, Denny T, Firrao G, Fischer-Le Saux M, Li X, et al. List of new names of plant pathogenic bacteria (2011-2012). Journal of Plant Pathology. 2014;96(2):223-6.

94. Pritchard L, Humphris S, Baeyen S, Maes M, Van Vaerenbergh J, Elphinstone J, et al. Draft genome sequences of four *Dickeya dianthicola* and four *Dickeya solani* strains. Genome Announcements. 2013;1(4).

95. Tian Y, Zhao Y, Yuan X, Yi J, Fan J, Xu Z, et al. *Dickeya fangzhongdai* sp. nov., a plant-pathogenic bacterium isolated from pear trees (*Pyrus pyrifolia*). International Journal of Systematic and Evolutionary Microbiology. 2016;66(8):2831-5.

96. Gardan L. Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and delineation of four novel species, *Dickeya*

dadantii sp. nov., *Dickeya dianthicola* sp. nov., *Dickeya dieffenbachiae* sp. nov. and *Dickeya zeae* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2005.

97. Denman S, Brady C, Kirk S, Cleenwerck I, Venter S, Coutinho T, et al. *Brenneria goodwinii* sp. nov., associated with acute oak decline in the UK. International Journal of Systematic and Evolutionary Microbiology. 2012;62(10):2451-6.

98. Brady C, Hunter G, Kirk S, Arnold D, Denman S. Description of *Brenneria roseae* sp. nov. and two subspecies, *Brenneria roseae* subspecies *roseae* ssp. nov and *Brenneria roseae* subspecies *americana* ssp. nov. isolated from symptomatic oak. Systematic and Applied Microbiology. 2014;37(6):396-401.

99. Hauben L, Moore ER, Vauterin L, Steenackers M, Mergaert J, Verdonck L, et al. Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. Systematic and Applied Microbiology. 1998;21(3):384-97.

100. Li Y, Xue H, Guo L-m, Koltay A, Palacio-Bielsa A, Chang J, et al. Elevation of three subspecies of *Lonsdalea quercina* to species level: *Lonsdalea britannica* sp. nov., *Lonsdalea iberica* sp. nov. and *Lonsdalea populi* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2017;67(11):4680-4.

101. Chari A, Oakeson KF, Enomoto S, Jackson DG, Fisher MA, Dale C. Phenotypic characterization of *Sodalis praecaptivus* sp. nov., a close non-insect-associated member of the *Sodalis*-allied lineage of insect endosymbionts. International Journal of Systematic and Evolutionary Microbiology. 2015;65(Pt 5):1400.

102. Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, et al. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect–bacterial symbioses. PLOS Genetics. 2012;8(11):e1002990.

103. Bercovier H, Steigerwalt AG, Guiyoule A, Huntley-Carter G, Brenner DJ. *Yersinia aldovae* (formerly *Yersinia enterocolitica*-like group X2): a new species of *Enterobacteriaceae* isolated from aquatic ecosystems. International Journal of Systematic and Evolutionary Microbiology. 1984;34(2):166-72.

104. Sprague LD, Neubauer H. *Yersinia aleksiciae* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2005;55(2):831-5.

105. Wauters G, Janssens M, Steigerwalt AG, Brenner DJ. *Yersinia mollaretii* sp. nov. and *Yersinia bercovieri* sp. nov., formerly called *Yersinia enterocolitica* biogroups 3A and 3B. International Journal of Systematic and Evolutionary Microbiology. 1988;38(4):424-9.

106. Neubauer H, Aleksic S, Hensel A, Finke E-J, Meyer H. *Yersinia enterocolitica* 16S rRNA gene types belong to the same genospecies but form three homology groups. International Journal of Medical Microbiology. 2000;290(1):61-4.

107. Batzilla J, Antonenka U, Höper D, Heesemann J, Rakin A. *Yersinia enterocolitica palearctica* serobiotype O: 3/4-a successful group of emerging zoonotic pathogens. BMC Genomics. 2011;12(1):348.

108. Neubauer H, Reischl U, Köstler J, Aleksic S, Finke E-J, Meyer H. Variations in the 165 rRNA gene sequence of *Yersinia enterocolitica* isolates influence the specificity of molecular identification systems. Zentralblatt für Bakteriologie. 1999;289(3):329-37.

109. Hurst MR, Becher SA, Young SD, Nelson TL, Glare TR. *Yersinia entomophaga* sp. nov., isolated from the New Zealand grass grub *Costelytra zealandica*. International Journal of Systematic and Evolutionary Microbiology. 2011;61(4):844-9.

110. Ursing J, Brennert JD, Bercovier H, Fanning GR, Steigerwalt AG, Brault J, et al. *Yersinia frederiksenii*: A new species of Enterobacteriaceae composed of rhamnose-positive strains

(formerly called atypical *Yersinia enterocolitica* or *Yersinia enterocolitica*-like). Current Microbiology. 1980;4(4):213-7.

111. Brenner DJ, Bercovier H, Ursing J, Alonso JM, Steigerwalt AG, Fanning GR, et al. *Yersinia intermedia*: A new species of enterobacteriaceae composed of rhamnose-positive, melibiose-positive, raffinose-positive strains (formerly called *Yersinia enterocolitica* or *Yersinia enterocolitica*-like). Current Microbiology. 1980;4(4):207-12.

112. Bercovier H, Ursing J, Brenner JD, Steigerwalt AG, Fanning GR, Carter GP, et al. *Yersinia kristensenii*: A new species of Enterobacteriaceae composed of sucrose-negative strains (formerly called atypical *Yersinia enterocolitica* or *Yersinia enterocolitica*-like). Current Microbiology. 1980;4(4):219-24.

113. Merhej V, Adékambi T, Pagnier I, Raoult D, Drancourt M. *Yersinia massiliensis* sp. nov., isolated from fresh water. International Journal of Systematic and Evolutionary Microbiology. 2008;58(4):779-84.

114. Murros-Kontiainen A, Fredriksson-Ahomaa M, Korkeala H, Johansson P, Rahkila R, Björkroth J. *Yersinia nurmii* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2011;61(10):2368-72.

115. Murros-Kontiainen A, Johansson P, Niskanen T, Fredriksson-Ahomaa M, Korkeala H, Björkroth J. *Yersinia pekkanenii* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2011;61(10):2363-7.

116. Niskanen T, Laukkanen R, Murros A, Björkroth J, Skurnik M, Korkeala H, et al. Characterisation of non-pathogenic *Yersinia pseudotuberculosis*-like strains isolated from food and environmental samples. International Journal of Food Microbiology. 2009;129(2):150-6.

117. Bercovier H, Mollaret H, Alonso JM, Brault J, Fanning GR, Steigerwalt AG, et al. Intra- and interspecies relatedness of *Yersinia pestis* by DNA hybridization and its relationship to *Yersinia pseudotuberculosis*. Current Microbiology. 1980;4(4):225-9.

118. Commission J. Opinion 60. Rejection of the name *Yersinia pseudotuberculosis* subsp. *pestis* (van Loghem) Bechovier et al. 1991 and conservation of the name *Yersinia pestis* (Lehman and Neumann) van Loghem 1944 for the plague bacillus. International Journal of Systematic Bacteriology. 1985;35:540-1.

119. Kim W, Song M-O, Song W, Kim K-J, Chung S-I, Choi C-S, et al. Comparison of 16S rDNA analysis and rep-PCR genomic fingerprinting for molecular identification of *Yersinia pseudotuberculosis*. Antonie Van Leeuwenhoek. 2003;83(2):125-33.

120. Eksić S, Steigerwalt AG, Bockemühl J, Huntley-Carter GP, Brenner DJ. *Yersinia rohdei* sp. nov. isolated from human and dog feces and surface water. International Journal of Systematic and Evolutionary Microbiology. 1987;37(4):327-32.

121. Ewing W, Ross A, Brenner DJ, Fanning G. *Yersinia ruckeri* sp. nov., the redmouth (RM) bacterium. International Journal of Systematic and Evolutionary Microbiology. 1978;28(1):37-44.

122. Janda JM, Abbott SL. The genus *Hafnia*: from soup to nuts. Clinical Microbiology Reviews. 2006;19(1):12-28.

123. Huys G, Cnockaert M, Abbott SL, Janda JM, Vandamme P. *Hafnia paralvei* sp. nov., formerly known as *Hafnia alvei* hybridization group 2. International Journal of Systematic and Evolutionary Microbiology. 2010;60(8):1725-8.

124. Ramos A, Dámaso D. Extraintestinal infection due to *Hafnia alvei*. European Journal of Clinical Microbiology and Infectious Diseases. 2000;19(9):708-10.

125. LESSEL EF. Status of the name *Proteus morganii* and designation of the neotype strain. International Journal of Systematic and Evolutionary Microbiology. 1971;21(1):55-7.

126. Jensen KT, Frederiksen W, Hickman-Brenner FW, Steigerwalt AG, Riddle CF, Brenner DJ. Recognition of *Morganella* Subspecies, with Proposal of *Morganella morganii* subsp. *morganii* subsp. nov. and *Morganella morganii* subsp. *sibonii* subsp. nov. International Journal of Systematic and Evolutionary Microbiology. 1992;42(4):613-20.

127. O'Hara CM, Brenner FW, Miller JM. Classification, identification, and clinical significance of *Proteus, Providencia*, and *Morganella*. Clinical Microbiology Reviews. 2000;13(4):534-46.

128. Tomczak H, Smuszkiewicz P. A rarely isolated micro-organism, *Budvicia aquatica,* cultured from urine of a patient with Guillain–Barré Syndrome. JMM Case Reports. 2014;1(2):e001503.

129. Ortega Á, Krell T. The HBM domain: introducing bimodularity to bacterial sensing. Protein Science. 2014;23(3):332-6.

130. Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiology and Molecular Biology Reviews. 1999;63(2):479-506.

131. Anantharaman V, Aravind L. Cache–a signaling domain common to animal Ca2+-channel subunits and a class of prokaryotic chemotaxis receptors. Trends in Biochemical Sciences. 2000;25(11):535-7.

Chapter 3

Elucidation of FliA and glucose regulation in *Dickeya dadantii* 3937 in minimal media

SUMMARY

Motility is a prominent characteristic of the majority of animal and plant pathogenic bacteria. Flagellar motility in combination with chemotaxis enables bacteria to explore their environments and locate their hosts. In *Escherichia coli*, flagella and chemotaxis related genes are controlled by the alternative sigma factor FliA, also known as σ^{28} or RpoF. FliA also regulates motility in *Dickeya dadantii* 3937. The presence or absence of nutrients has an effect on virulence. Notably, glucose suppression of virulence genes is well-known in *E. coli*. We conducted RNA-sequencing analysis of wild-type *D. dadantii* 3937 and a *fliA* mutant in minimal media with either 0.2% glycerol or 0.2% glucose to determine the genes regulated by FliA and glucose. The *D. dadantii* flagellar genes are regulated by FliA as expected based on flagellar gene regulation studies in *E. coli.* Even though 12 of 47 mcp have a canonical FliA binding site upstream, none of them were regulated by FliA in the conditions tested. We expected virulence genes to be repressed in minimal medium containing glucose, but several were upregulated compared to cells grown in glycerol, including the butanediol production genes (*budA, budB, budC,* and *budR*), chrysobactin genes and transport (*cbsA, cbsB, cbsC, cbsE, cbsF, cbsH, cbsP,* and *fct*), metalloproteases (*prtA, prtB,* and *prtC*), the type III secretion system pilus HrpA and T3SS-secreted HrpN, avirulence proteins AvrL and AvrM, capsular polysaccharide biosynthesis proteins, expansin, and the plant antimicrobial peptide resistance protein YbjX. Overall, our results show that as yet to be identified sigma factor regulates the 47 *D. dadantii mcp*, that

several known virulence genes are not repressed by carbon catabolite repression in *D. dadantii*, and that only a subset of genes suspected to be regulated by PecS are induced by glycerol.

INTRODUCTION

Flagellar motility is important for pathogenesis of *Dickeya* and *Pectobacterium*, two related genera of plant pathogenic bacteria that cause soft rot, wilt, and blackleg disease in numerous angiosperm plant species (1, 2). Motility may play a role in finding plant roots, wounds through which bacteria can enter plants (3, 4), or may aid bacteria in moving in disease lesions or through xylem (5, 6). Like other members in the order Enterobacteriales, *Dickeya* encodes peritrichous flagella (2). The bacterial flagella are secreted through a flagellar type III secretion system (T3SS) and their rotation is powered through ATP hydrolysis (7).

These bacteria swim (taxis) toward attractants through the action of the chemotaxis system. The direction of flagella rotation is controlled by the chemotaxis system, which consists of methyl-accepting chemotaxis proteins (Mcp) and the chemotaxis (Che) proteins. The default rotation of flagella is counter-clockwise (CCW), in which bacteria swim in a straight direction (run) (8). Mcp, also known as chemoreceptors, perceive environmental and intracellular signals through the ligand binding domain (9). Upon ligand binding, Mcp undergo conformational change that is relayed through the transmembrane domain to the cytoplasmic domain (9). The cytoplasmic domain interacts with the histidine autokinase CheA and the scaffolding protein CheW (10). The autophosphorylation activity of CheA is controlled by the Mcp cytoplasmic domain based on the signals sensed (10, 11). The presence of attractants inhibits CheA autophosphorylation (OFF state) and vice versa (ON state) (10, 11). Phosphorylated CheA performs kinase activity on CheY, which binds to the flagellar motor switch, inducing a clockwise (CW) rotation of the flagella, resulting in a change of bacterial swimming direction

(tumble) (8). Phosphorylated CheY interaction with the flagellar motor switch is brief due to dephosphorylation by the phosphatase CheZ (10). Phosphorylated CheB and CheR are members of the chemotaxis adaptation pathway that balances the ON-OFF state through demethylation and methylation of the Mcp cytoplasmic domain, respectively (10). Mcp, which are accessory proteins for the flagellar system, are somewhat analogous to T3SS effectors. Both T3SS effectors and Mcp are modular. The T3SS effectors require sigma factor HrpL for expression, which is the same extracytoplasmic function (ECF) family sigma factor required for T3SS gene expression. In *Escherichia coli*, the *mcp* require FliA for expression and FliA is an ECF sigma factor required for expression of several flagellar secretion system genes. The Mcp differ, however, in that they are not secreted through a T3SS and, unlike effectors, they do not manipulate host cells.

The flagellar loci and associated regulatory genes are well-studied in the bacterial model organism of the order Enterobacteriales, *E. coli*. In *E. coli*, flagellar genes are expressed in three different tiers, or classes, based on temporal mode of expression (12). The sole class I operon, *flhDC*, encodes the master flagellar transcriptional complex FlhDC, which activates all related components of the bacterial flagellar system (12). FlhD and FlhC are unique among regulatory genes in the Enterobacteriales genome and there are no other paralogs of these genes in Enterobacteriales genomes. FlhDC directly activates promoters of class II operons (*flgAMN*, *flgBCDEFGHIJ*, *flhBAE, fliAZY, fliE, fliFGHIJK,* and *fliLMNOPQR*), which are transcribed by the holoenzyme complex of RNA polymerase (RNAP) and the primary sigma factor 70 (σ^{70}) (13). The class II operons code for proteins involved in the biosynthesis of the flagellar base and hook structures (including their corresponding T3SS export components), and the two antagonistic regulators, the alternative sigma factor FliA (σ^{28}) and its anti-sigma factor FlgM (13). FlgM interacts with FliA in the cytoplasm, which prevents the sigma factor from interacting

with RNAP, thus inhibiting the transcription of the class III operons that encode proteins involved with late flagellar assembly (12). FlgM is secreted out of the cytoplasm upon the completion of the flagellar hook-basal structure, allowing the transcription of the class III operons (*flgKL, flgMN, fliC, fliDST, motAB-cheAW,* and *tar-tap-cheRBYZ*) by the FliA holoenzyme (12, 13). FliA also regulates the transcription of the class II operons *fliAZY* and *fliLMNOPQR* (14). All of these regulatory genes are present in *Dickeya* and *Pectobacterium* (Figure 3.1).

The *E. coli* FliA binding site is TAAAGTTT-N_{11}-GCCGATAA (15). In *E. coli*, FliA also regulates the chemotaxis system and there are FliA binding sites upstream of all of the *E. coli* Mcp genes, *tar*, *tap*, *trg*, *tsr*, and *aer* (12). Unlike *E. coli*, which tends to encode 3 to 5 *mcp*, the plant pathogens *Dickeya* and *Pectobacterium* species encode at least 30 *mcp* and as many as 47 *mcp* (3, 16). This suggests that these plant pathogens encounter and respond to a wider range of stimuli while swimming or swarming than related animal pathogens. Known stimuli sensed by *Dickeya* and *Pectobacterium* Mcp include D-sugars (fructose, galactose, glucose, maltose, mannose, ribose, sucrose, and xylose), L-amino acids (alanine, aspartate, glutamine, glycine, proline, serine, tryptophan), organic acids (citric acid, glutamic acid, malic acid, and succinic acid), glycerol, mannitol, oligogalacturonides, and the plant hormone jasmonic acid (3, 17, 18).

Gene expression experiments in *Dickeya dadantii* 3937 showed that not all *mcp* are expressed in culture with minimal media (19, 20). This is expected since constitutive expression of all 47 *mcp* is unlikely and instead, expression is likely to respond to environmental cues, but also to be co-regulated with flagella, in at least some cases. For example, in *E. coli*, CRP, which regulates bacterial genes in response to glucose levels, represses motility when glucose levels are high (21). Furthermore, *E. coli* virulence genes were suppressed in the presence of glucose (22).

In *Dickeya,* pectate lyase genes are suppressed by PecS, a horizontally acquired transcriptional regulator that is repressed when glucose levels are high (1). In addition, biofilm formation in *Dickeya* is induced when glycerol is present (23). However, FliA and glucose regulation of *Dickeya* and *Pectobacterium* remain mostly unexplored. In this work, our goal was to determine the regulon of FliA and glucose in *D. dadantii* 3937.

MATERIALS AND METHODS

Bacterial strains and growth conditions

A colony of wild-type (WT) *D. dadantii* strain 3937 or a *fliA* mutant (24) were grown to early log phase in M9 minimal medium (25) containing 0.2% glucose (MM-glu) or 0.2% glycerol (MM-gly) as carbon sources at 30°C with shaking at 220 rpm. The early log phase was determined based on the growth curves of the strains in each medium (data not shown). For WT strain in MM-gly, the OD_{600} ranged from 0.10 to 0.69 (approximately $8.3*10^7$ cfu/ml to $6.4*10^8$ cfu/ml). In MM-glu, the OD_{600} for the WT strain ranged from 0.54 to 0.68 (approximately $3.7*10^8$ cfu/ml). For the *fliA* mutant strain, the OD_{600} in MM-gly ranged from 0.10 to 0.80 (approximately $1.2*10^8$ cfu/ml to $9.5*10^8$ cfu/ml), while the OD_{600} in MM-glu ranged from 0.27 to 0.41 (approximately $1.7*10^8$ cfu/ml to $5.0*10^8$ cfu/ml). Four biological replicates were prepared for each growth condition to acquire a total of 16 samples. For each sample, cells from 12 mL of culture were harvested by centrifugation at 5000 x g at 4°C for 2 min. The pellets were treated with 20 ul RNA*later*® (Sigma-Aldrich, USA) for RNAse inactivation.

Total RNA isolation, rRNA depletion and quality control

Total RNA from each treated cell pellets was isolated using the TRIzol™ Plus RNA Purification Kit (Invitrogen; Thermo Fisher Scientific, USA) following the manufacturer's protocol. The total RNA was further purified using RNA Clean and Concentrator™-5 kit (Zymo

Research, USA) with DNAse treatment according to the manufacturer's user guide. The quantity and quality of the total RNA was determined using Qubit 3.0 fluorometer (Invitrogen; Life Technologies; Thermo Fisher Scientific, USA) and Agilent Technologies 2200 TapeStation system (Agilent Technologies, USA), respectively. The RNA integrity number (RIN) of the samples ranged from 8.2 to 8.7, indicating high quality of RNA based on previous study in *D. dadantii* 3937 in which the recommended threshold is above 7.0 (26). Ten ug of total RNA from each sample was subjected to the RiboMinus™ Transcriptome Isolation Kit (Invitrogen; Thermo Fisher Scientific, USA) for the removal of rRNA. The rRNA-depleted samples were purified using RNA Clean and Concentrator™-5 kit. Their concentrations were measured by Qubit 3.0 fluorometer, and the integrity of each sample was assessed with an Agilent TapeStation 2200. RIN values ranges from 2.3 to 5.1, indicating the successful depletion of rRNA.

Construction of cDNA library and Illumina sequencing

cDNA Library was constructed using KAPA RNA HyperPrep Kit for Illumina platforms (KAPA Biosystems, USA). Fragmentation was performed on 100 ng RNA samples by heating at 94°C for 6 min to obtain a mean library insert size of 200 - 300 bp. First-strand cDNA synthesis was carried out with the following conditions: 25°C for 10 min, 42°C for 15 min, and 70°C for 15 min. Second-strand cDNA synthesis was performed concurrently with A-tailing under the conditions: 16°C for 30 min and 62°C for 10 min. Adapter ligation was completed at 20°C for 15 min using KAPA Dual-Indexed Adapter Kit for Illumina platforms (KAPA Biosystems, USA). The adapter-ligated DNA was purified twice with KAPA Pure Beads, followed by PCR amplification using the following thermocycling profile: 98°C for 45 sec; 8 cycles of 98°C for 15 sec, 60°C for 30 sec, and 72°C for 30 sec; and 72°C for 1 min. The amplified library DNA was purified with KAPA Pure Beads.

Sequencing was carried out at the Colorado State University next-generation sequencing core facility (Colorado, USA; https://www.research.colostate.edu/genomics/). The quality and quantity of the libraries were determined using Agilent TapeStation 2200, and Qubit 2.0 fluorometer. Libraries were sequenced using NextSeq 500/550 High Output Kit (Illumina, USA) on Illumina NextSeq^{TM} 500 platform, generating 75 bp single-end reads.

RNA sequencing data analysis

The obtained unpaired reads were quality-checked using FASTQC (27). Subsequently, the low-quality reads and the adaptors were removed via TRIMMOMATIC (version 0.36) (28). The trimmed reads were mapped to the *D. dadantii* 3937 genome (https://www.ncbi.nlm.nih.gov/nuccore/NC_014500.1?report=fasta) using HISAT2 (version 2.1.0) (29). The number of reads per gene was tabulated using featureCounts (version 2.0.0) (30). Differentially expressed genes (DEG) were determined using a log2 fold change cut-off of 1 and -1, and a False Discovery Rate (FDR) threshold of 5% (adjusted p-value < 0.05) in R software (31), utilizing the DESeq2 package (version 1.26.0) (32). Volcano plots were made using the web app VolcaNoseR (33).

Identification of sigma factors and flagellar gene and operons

For identification of orthologous sigma factors in Pectobacteriaceae family, protein sequences of *Escherichia coli* ATCC 11775 sigma factors were retrieved from the ASAP database (34, 35) and BLASTP search was conducted against the type strains in the plant pathogenic genera *Brenneria, Dickeya, Lonsdalea,* and *Pectobacterium*. In order to determine whether the flagellar gene organization and the operon structure are conserved across different bacterial species in the order Enterobacteriales, manual identification was carried out using the genomes of the strains (listed in Table 3.1) available in the ASAP database (34, 35).

Identification of FliA binding sites upstream of flagellar gene/operons and methyl-accepting chemotaxis protein genes of *D. dadantii* 3937

In *E. coli,* FliA binding sites were found upstream of the flagellar gene/operons and chemotaxis genes. Nucleotide sequences of 500 bp upstream of five *D. dadantii* 3937 flagellar gene/operons and 47 methyl-accepting chemotaxis protein genes were retrieved from the ASAP database (34, 35) for manual identification of putative FliA binding site in these upstream regions.

RESULTS

Conservation of flagellar operons and FliA binding sites in the Pectobacteriaceae

The flagellar operons are conserved across the Pectobacteriaceae and the operon structure is similar to other members of the Enterobacteriales, with a few exceptions (Figure 3.1). Plant pathogenic genera *Dickeya, Brenneria, Pectobacterium,* and *Lonsdalea* have one *mcp* gene upstream of the *cheRBYZ* operon. Whereas in animal pathogens *Escherichia, Salmonella, Hafnia, Morganella,* and *Yersinia,* and the insect symbiont *Sodalis*, *mcp* gene is part of the *cheRBYZ* operon. *Sodalis, Morganella,* and *Yersinia* have additional copies of the flagellin gene, *fliC.* The *fliDST, fliE, fliFGHIJK,* and *fliLMNOPQR* operons are in the opposite direction in *Dickeya paradisiaca* and *Lonsdalea. Dickeya aquatica, Dickeya paradisiaca, Sodalis,* and *Morganella* are missing *flhE.*

The FliA binding sites are also conserved in the promoters of the class III flagellar genes in *D. dadantii* 3937 and are similar to those found in *E. coli* (Table 3.2) (12). The two conserved *che* operons (*motAB-cheAW* and *tar-tap-cheRBYZ*) have a FliA binding site in both *E. coli* and *D. dadantii* (Table 3.2 & 3.3). We examined 47 methyl-accepting chemotaxis genes (*mcp*) from

D. dadantii 3937 and found that 12 *mcp* had putative FliA binding sites in their promoters (Table 3.3), suggesting that only a subset of *mcp* could be directly regulated by FliA.

RNA-sequencing

RNA-seq of 16 samples generated approximately 361 million unpaired reads with an average of 22 million reads per library. After trimming, about 354 million surviving reads (~98%) were obtained. A total of approximately 350 million reads were mapped with an average of 98% of mapped reads per library.

FliA regulates *D. dadantii* flagellar genes

Growing the *fliA* mutant of *D. dadantii* 3937 in MM-glu resulted in a differential regulation of 23 genes when compared to the wild type. Of these, 7 genes were up-regulated, and 14 genes were down-regulated (Figure 3.2 & Table 3.4). We observed that 5 up-regulated genes (*flgA, flgB, fliE, fliF, fliZ*) and 8 down-regulated genes (*fliC, fliD, fliS, fliT, flgK, flgL, flgM,* and *flgN*) were flagellar genes, all of which were regulated as expected in a *fliA* mutant based on work in *E. coli* (12). FliA also represses a conjugal transfer protein (*traF*) homolog (ABF-0016857). In this case, the regulation is likely indirect since there is no apparent FliA consensus binding site in the *traF* upstream regions. Additionally, FliA represses *gpV*, a gene encoding the baseplate protein for type VI secretion system (T6SS). FliA is required for up-regulation of two *mcp* (ABG-0017097 and ABF-0017665), but neither has a consensus FliA binding site. In addition, FliA is required for up-regulation of four hypothetical proteins (ABF-0014838 (*flxA*); ABF-0016874; ABF-0017664 (*tolA*); and ABF-0018601). Of these, three of the genes, ABF-0014838, ABF-0016874, and ABF-0018601, have a FliA consensus binding site upstream (Table 3.2). Both ABF-0014838 (*flxA*) and ABF-0018601 genes are found primarily in a subset of *Dickeya* and *Pectobacterium* species and both encode small glutamine rich proteins.

In the glycerol-based medium (MM-gly), only one flagellar gene, *fliZ*, was up-regulated in the *fliA* mutant (Figure 3.3 & Table 3.4). The *traF* homolog was also up-regulated under this condition. A total of seven flagellar genes were down-regulated in the *fliA* mutant compared to the WT when the cells were grown in MM-gly, instead of eight flagellar genes as in MM-glu (*flgL* was not differentially expressed in MM-gly). Three of the four hypothetical proteins induced by FliA in MM-glu were also induced in MM-gly (ABF-0014838 (*flxA*); ABF-0016874; and ABF-0017664 (*tolA*)). In addition, a putative motility protein YjfB (ABF-0018207) and two *mcp* (ABF-0019050 and ABF-0020167) were induced by FliA in MM-gly. Both YjfB and the *mcp* (ABF-0019050) have a FliA binding motif upstream (Table 3.2 & 3.3). All genes are either the first in the operon or in single gene operons.

FliA and *D. dadantii mcp*

Although putative FliA consensus binding sites were found upstream of 12 *D. dadantii mcp*, only one (ABF-0019050) was significantly down-regulated in *fliA* mutant in MM-gly (Figure 3.3 & Table 3.4). The transcript level of ABF-0019050 in WT cells was also lower in MM-glu compared to MM-gly, suggesting that this gene is suppressed by glucose, although the downregulation was not significant (Table 3.5). We found that many other genes that encode chemotaxis-related proteins were differentially expressed in WT *D. dadantii* 3937 in MM-glu compared to MM-gly, despite not being significant (Table 3.5). For example, 13 *mcp* (including two aerotaxis genes *aerA* and *aerC*) had lower transcript levels in MM-glu and 34 *mcp* (33 *mcp* and an aerotaxis gene *aerB*) had higher transcript levels in MM-glu (Table 3.5). In addition, the average normalized transcripts of chemotaxis proteins (CheA, CheB, CheR, CheV, CheW, and CheZ) were high, ranges from 2100 to 8015, in MM-glu (Table 3.5).

Three *mcp* (ABF-0015513; ABF-0018754; and ABF-0019858) were among the 500 most-expressed genes (more than 1,500 average normalized counts) in *fliA* mutant compared to WT *D. dadantii* 3937 in both MM-glu and MM-gly (data not shown) and an additional nine *mcp* were in the next 500 most expressed genes (ABF-0016115; ABF-0016979; ABF-0018502; ABF-0018541; ABF-0018765; ABF-0019309; ABF-0019790; ABF-0019851; and ABF-0020431), demonstrating that mcp were not generally repressed, but instead that they were not differentially regulated when *fliA* was deleted. Of these genes, four *mcp* (ABF-0016115; ABF-0018541; ABF-0018754; and ABF-0018765) have a FliA consensus binding site upstream (Table 3.3). Similarly, the chemotaxis protein genes *cheA, cheW, cheR, cheB, cheY,* and *cheZ,* together with the flagellar motor genes *motA* and *motB,* were not differentially regulated when *fliA* was deleted in both MM-glu and MM-gly (data not shown). These genes are regulated by FliA in *E. coli* (12). We would expect *mcp* to be co-regulated with the *che* genes and flagellar motor protein genes based on work in *E. coli.* However, even with a strong FliA consensus site, there is little evidence of high differential expression, suggesting that *mcp, che* genes and flagellar motor genes are regulated by a different sigma factor.

Several *mcp* appeared to be well-expressed in the tested conditions, but to not require FliA for expression. Sigma factors are required for initiation of gene transcription in bacteria and therefore a different sigma factor must be responsible for *mcp* expression. We examined sigma factors encoded in plant pathogenic Pectobacteriaceae species (*Brenneria, Dickeya, Lonsdalea,* and *Pectobacterium*) and found that they encode six conserved sigma factors: RpoD (σ^{70}), RpoE (σ^{24}), RpoH (σ^{32}), RpoN (σ^{54}), RpoS (σ^{38}), and FliA (RpoF/σ^{28}) (Table 3.6). Most species also encode HrpL, which activates the Hrp type III secretion system (T3SS). At least 12 additional sigma factors are encoded by plant pathogens in this family (data not shown), but their presence

is sporadic, so these sigma factors are unlikely to be important for *mcp* expression. Most of the additional sigma factors are homologous to FecI, which regulates genes required for iron uptake in *E. coli* (36). Therefore, there is currently no clear candidate for a sigma factor likely to initiate expression of the *mcp* in *D. dadantii.*

***D. dadantii* up-regulated 224 genes and down-regulated 103 genes when grown in glucose compared to glycerol**

Several genes important for plant-microbe interactions were up-regulated in MM-glu, including *budA, budB, budC,* and *budR* that are involved in butanediol production (37, 38), chrysobactin-related genes *cbsA, cbsB, cbsC, cbsE, cbsF, cbsH, cbsP,* and *fct* for iron acquisition (39, 40), TonB-dependent receptor FusA for iron acquisition from plant ferredoxin (41), the type I secreted metalloproteases-encoding genes *prtA, prtB,* and *prtC* (42, 43), the T3SS *hrpA* pilus, and *hrpN* hairpin required for cell aggregation (44, 45), avirulence protein genes *avrL* and *avrM* (46), capsular polysaccharide biosynthesis proteins, and colanic acid synthesis enzymes WcaJ and WcaB (47), expansin gene *yoaJ* (48), xylanase (*xynA*), and plant antimicrobial peptide resistance protein YbjX (49) (Figure 3.4 & Table 3.7). This was unexpected because pectate lyases, which are the main pathogenicity factor of *Dickeya* and *Pectobacterium*, are repressed by the cyclic-AMP receptor protein (CRP) system when glucose is present (1). Additionally, genes involved in microbe-microbe interactions were upregulated in glucose, namely the T6SS haemolysin-coregulated proteins (Hcp) and valine-glycine repeat proteins (Vgr) (50). Furthermore, the endotoxin genes *cytA* and *cytD*, which are homologous to *Bacillus thuringiensis* Cyt toxin gene, and the toxin-antitoxin systems YafN/YafO were upregulated in glucose (51).

In addition to these virulence genes, motility-related genes were up-regulated in MM-glu compared to MM-gly, which include surfactant encoded gene *rhlA* for swarming motility (20),

flagellar gene *flhE*, and five *mcp*. Two *mcp* have the four-helix bundle (4HB) ligand binding domain (ABF-0017665 and ABF-0017674). While ABF-0016380 and ABF-0017537 have the Cache (Ca^{2+} channels and chemotaxis receptors) domain, and ABF-0015600 has the helical bimodular (HBM) domain. Two of the five *mcp* are in the same locus (ABF-0017665 and ABF-0017674). Of these five, only one has potential FliA binding site upstream (ABF-0016380) (Table 3.3). A total of 35 ABC transporter genes of wild-type *D. dadantii* were up-regulated in MM-glu compared to MM-gly. These ABC transporter genes fall into 15 transporters that belong to 6 families (i.e., family 2, 4, 5, 6, 8, 9) and are likely to import peptides, amino acids, metal (heme, ferric chrysobactin and zinc) and nitrate, and export antimicrobials.

Of the 103 down-regulated genes in MM-glu, nine were structural genes in the flagella, showing that the presence of glucose suppresses flagellar motility through FlhDC inhibition by CRP (Figure 3.4 & Table 3.8). This finding is congruent with *E. coli* (12). The T3SS *hrpE* and the adhesin transporter, *vfmX* were down-regulated in MM-glu compared to MM-gly. Only one of 47 *mcp* was down-regulated in MM-glu. This mcp (ABF-0014618) does not have a FliA binding site. Genes involved with regulation were also down-regulated in MM-glu, namely *pecS, rpoS, xylR, rhaR,* and *rhaS*. The repression of the stationary sigma factor *rpoS* was expected because CRP regulates *rpoS* expression in *E. coli* (52). The downregulation of the putative xylanase repressor (*xylR*) explains the upregulation of xylanase (*xynA*) in MM-glu. Furthermore, genes involved with metabolism of carbon sources besides glucose were down-regulated in MM-glu compared to MM-gly, such as *glpABC, glpD, glpK, glpQ, rhaD, ganABC, iolA, iolC, iolD,* and *sfcA/maeA*. Super oxide dismutase (*sodC*) and *atpF* (F0F1 ATP synthase subunit B), which are responsible for destroying reactive oxygen species, a defense mechanism of plant cells, were also down-regulated in MM-glu.

In addition, 23 ABC transporter genes that fall into 10 transporters were down-regulated in MM-glu. Of these 10 transporters, six are putative sugar transporters (three in family 1 and three in family 5), three are peptide transporters, and one is amino acid transporter. Genes encoding transporters and metabolism genes for sugars other than glucose are typically down-regulated in the presence of glucose by the CRP system, so this was an expected result. In contrast, however, when glycerol is present as a carbon source, all these sugar transporters are up-regulated. This represents 86% of the total sugar transporters encoded by *D. dadantii*.

ABC transporters

In total, 55 ABC transporter genes of wild-type *D. dadantii* that belong to nine families were differentially regulated in MM-glu compared to MM-gly (Table 3.9). They are sugar transporters (family 1 and 5), peptide transporters (family 2 and 10), amino acid transporters (family 4 and 5), metal transporters (family 8), putative transporters (family 3; ABF-0016441), and exporters (family 6 and 9).

In the presence of glucose, genes encoding a sugar transporter for the uptake of maltose and maltodextrin was up-regulated but eight genes encoding six other transporters for importing arabinose, fructose, xylose, ribose, galactan, galacofuranose, and pectic oligomers were down-regulated.

A comparative analysis of ABC transporters identified that the genome of *D. dadantii* encodes 10 ABC transporters for peptide uptake (data not shown) and of these, two was up-regulated in MM-glu (ABF-0019444 and ABF-0019568) and two was up-regulated (DppABCDF and OppABCDF) in MM-gly. In addition, *D. dadantii* encodes two carbon starvation proteins, which may import peptides under starvation conditions, and one was up-regulated in MM-gly (ABF-0019249) and the other in MM-glu (*yjiY*/*cstA*; ABF-0019365).

Interestingly, 11 genes encoding four amino acid transporters were up-regulated in MM-glu. Three transporters, including HisJMPQ, YhdWXYZ, and a putative transporter (ABF-0018788), are in family 4, while PotABCD for spermidine or putrescine uptake belongs to family 5. Without the presence of glucose, *D. dadantii* down-regulated genes encoding a branched-chain amino acid transporter, LivFGHKM.

Importing metal is crucial for bacterial survival. When *D. dadantii* grew in glucose, upregulation of six genes that encode four ABC metal ion transporters, including heme (HmuSTUV), ferric chrysobactin (CbuBCDG), iron-sulphur (FusCD), and zinc uptake (ABF-0019222 and ABF-0019223), were observed. These data suggest that the presence of glucose stimulates metal ion metabolism in *D. dadantii*.

In addition to ABC importers being differentially regulated, we observed *macB*, a gene consists of an ABC exporter MacAB was up-regulated in MM-glu compared to MM-gly. This transporter is known to export large protein toxins and virulence factors that contribute to antimicrobial resistance.

Protein secretion systems and capsule biosynthesis

Although the Hrp T3SS and Prt T1SS were not differentially regulated, proteins secreted through these systems were up-regulated in MM-glu compared to MM-gly (Figure 3.4 & Table 3.7). *D. dadantii* encodes 14 genes that contribute to capsule biosynthesis and nine of these 14 genes were up-regulated in MM-glu. In contrast, genes in the type III flagellar secretion system operons were down-regulated in MM-glu compared to MM-gly (Figure 3.4 & Table 3.8). This suggests that motility is regulated in opposition to genes that contribute to biofilm formation, which is expected, and that glucose or glycerol may play a role in this signaling. In addition, six *mcp* were differentially regulated, with five up-regulated in MM-glu and one in MM-gly. In

addition, *avrM* and *avrL* (a homolog of *Xanthomonas campestris* avirulence gene) were up-regulated in MM-glu (Figure 3.4 & Table 3.7). The function of these avirulence genes in *D. dadantii* is currently unknown.

Furthermore, four of five haemolysin-coregulated protein (Hcp) genes were up-regulated in MM-glu compared to MM-gly. Hcp are part of the type VI secretion systems (T6SS) apparatus, forming a filamentous tube, which is capped by valine-glycine repeat proteins (VgrG) (50). Two of three VgrG genes were also up-regulated in MM-glu. Despite being secreted through the T6SS, the Rhs (rearrangement hotspot) toxins involved in bacterial intercellular competition were not differentially regulated in MM-glu. These findings are in contrast with the T3SS and T1SS, where the proteins secreted through T6SS were not differentially regulated in MM-glu compared to MM-gly.

Secondary metabolites

Some secondary metabolites important for virulence were up-regulated in MM-glu compared to MM-gly. These include *budA, budB, budC*, and *budR* which are required to produce a volatile molecule, 2,3-butanediol. Both *Dickeya* and *Pectobacterium* produce copious amount of butanediol during infection, which increase plant apoplast pH to promote pectate lyase activity (37, 38). Additionally, butanediol serves as a signaling molecule that triggers the systemic acquired resistance in host plants (38). The *bud* genes upregulation in MM-glu suggests that glucose might be a signal that triggers their expression.

Glycerol metabolism

Glycerol is an alternative carbon source that can be utilized by bacteria. The mechanisms of glycerol transport and metabolism have been well-studied in *E. coli*. Glycerol is transported across bacterial cytoplasmic membrane via facilitated diffusion through the aquaglyceroporin

GlpF (Figure 3.5) (53). Upon entry into the cytoplasm, glycerol is phosphorylated to glycerol-3-phosphate (Gly3P) by the glycerol kinase GlpK (54). Gly3P can also be imported into the cytoplasm by the antiporter GlpT in a coupled export of inorganic phosphates (P_i) to the periplasm (55). The periplasmic Gly3P is a by-product of the hydrolyzation of deacylated phospholipids by the glycerophosphoryl diester phosphodiesterase GlpQ (56, 57). In the cytoplasm, Gly3P undergoes oxidization reaction catalyzed by either the aerobic dehydrogenase GlpD or the anaerobic dehydrogenase GlpACB, depending on the availability of oxygen (58). The end product of the reaction, glycerone phosphate/dihydroxyacetone phosphate (DHAP) is further metabolized in glycolysis. In this study, we found eight genes that are involved in glycerol metabolism (*glpF, glpK, glpT, glpQ, glpD,* and *glpACB*) were down-regulated in MM-glu compared to MM-gly in the WT *D. dadantii* 3937, which is expected since these genes are regulated by CRP in *E. coli* (Figure 3.4 & Table 3.8) (52).

Gene regulation by the Cyclic-AMP receptor protein

Transcription activation by the cyclic-AMP receptor protein (CRP) and its allosteric effector, cyclic-AMP (cAMP) has been studied for decades in *Escherichia coli* (52, 59, 60). CRP-cAMP complex binds to specific DNA sites to enhance the initiation of transcription by RNA polymerase holoenzyme (RNAP). The second messenger cAMP is synthesized by adenylyl cyclase when the concentration of glucose is relatively low (61). Under excess glucose condition, intracellular level of cAMP is low, resulting in carbon catabolite repression of genes involved in metabolism of other carbon sources besides glucose (62). Upon manual search in the upstream region of genes/operons that were down-regulated in MM-glu compared to MM-gly, we found putative CRP binding sites upstream of 21 genes (Table 3.10). Among these genes, one gene (*cstA*) was up-regulated in MM-glu compared to MM-gly. Notably, the CRP binding sites

upstream of the *glpFK, glpABC,* and *glpTQ* operons and *glpD* in *D. dadantii* 3937 are similar to the findings by Larson et. al, 1992 in *E. coli* K-12 (63).

PecS is up-regulated in MM-gly

PecS is a regulator encoded on a large island of horizontally acquired virulence genes present in *Dickeya* but not in other Pectobacteriaceae. Deletion of *pecS* has pleiotropic effects on *D. dadantii*, causing the cells to grow more slowly and this deletion impacts the expression of more than 600 genes (64). Some of these genes may be affected due to the pleiotropic nature of this mutation and not due to direct regulation by PecS. Glycerol affects *pecS* expression and we found that *pecS* was up-regulated in MM-gly compared to MM-glu. We compared genes listed as potentially regulated by *pecS* (*pecS* mutant vs WT at 6h post-inoculation) with genes up-regulated in MM-glu compared to MM-gly in WT *D. dadantii* 3937. Forty-two (42) out of the 575 genes reported to be *pecS* regulated at 6h post-inoculation in *Arabidopsis* were also regulated in the same manner directly when WT cells were grown in MM-glu (Table 3.11). Of these, 36 appeared to be repressed by PecS and six were induced by PecS. Of the remaining 533 genes differentially regulated at 6h post-inoculation, 11 that were differentially regulated in MM-glu compared to MM-gly were regulated in the opposite direction in this study. Overall, 522 remaining genes that were differentially regulated in *pecS* mutant at 6h post-inoculation in *Arabidopsis*, were not identified in our experiment. Together, these data suggest that PecS may affect regulation of 53 genes in minimal media and that it acts primarily as a repressor.

DISCUSSION

Dickeya and *Pectobacterium* strains each encode 30 or more *mcp*, suggesting that they can recognize and taxis toward multiple types of chemical stimuli. In *D. dadantii*, 12 of the *mcp* are preceded by a typical FliA binding site, suggesting that FliA is required for expression of at

least some *mcp*. We used RNA-seq to identify the FliA regulon and found that FliA regulates flagellar operons as expected. We also found that FliA regulates a conjugal transfer protein, a type VI secretion protein, and hypothetical proteins. Of these genes, five have a consensus FliA binding site (ABF-0014838 (*flxA*), ABF-0016874, ABF-0017664 (*tolA*), ABF-0018207 (*yjfB*), and ABF-0018601) (Table 3.2). Four *mcp* were down-regulated in *fliA* mutant background, but only one *mcp* has a putative FliA binding site (ABF-0019050). None of the remaining 43 *mcp* were differentially regulated by FliA under the tested conditions. The average normalized counts from the RNA seq experiments demonstrated that many of the *mcp* were expressed. For example, 12 of the mcp were among the 25% of genes with the highest mRNA counts in *fliA* mutant background. This data suggests that one or more of the nine other sigma factors besides FliA in *D. dadantii* are required for expression of these *mcp*. Of these, HrpL is unlikely to be involved since its regulon has been examined closely (65). RpoS regulates motility genes in *E. coli* via repression of FliA (66) and can regulate genes during exponential phase growth, so it is a potential candidate for *mcp* gene expressions. Another potential candidate would be the nitrogen utilization sigma factor RpoN, which activates the transcription of *flhDC* in *E. coli* (66). A biological rationale for the lack of co-regulation of flagella and the numerous *mcp* remains elusive.

In addition, we discovered that numerous genes involved in plant-microbe interactions are not under catabolite repression. As expected, PecS, one of the global regulators for virulence genes in *Dickeya*, was upregulated in MM-gly (20). The PecS regulon include pectate lyases, polygalacturonases, T3SS hairpin HrpN, and indigoidine (20). The upregulation of PecS in MM-gly provides further support that glycerol may be considered a signaling molecule for bacterial pathogenicity. In *Arabidopsis thaliana*, exogenous glycerol treatment increased the endogenous

production of glycerol-3-phosphate, a signal metabolite for systemic acquired resistance in plants (67-69). Furthermore, hydrogen peroxide (reactive oxygen species) level was increased as well, which could explain the upregulation of *sodC* and *atpF* in MM-gly.

Surprisingly, our study revealed that FliA is not required for the transcription of methyl-accepting chemotaxis genes in *Dickeya dadantii* in minimal media and the sigma factor required for initiation of expression of this large cadre of genes remains unknown. The downregulation of the *che* and *mot* genes in *fliA* mutant background validates FliA regulation of these genes as found in *E. coli.* However, the lack of differential expression points out to the possibility of repression of these genes by an unknown factor. Surprisingly, we found that several virulence genes were induced when glucose is the carbon source, suggesting that they are not controlled by carbon catabolite repression via CRP. Finally, our data support previous work with PecS that shows that glycerol could be a signal molecule perceived by bacteria for induction of some virulence genes.

Table 3.1: List of strains used in determination of flagellar gene organization and operon structure conservation in the order Enterobacteriales.

Family	**Genus**	**Strain**
Pectobacteriaceae	*Brenneria*	*Brenneria goodwinii* FRB 141
		Brenneria roseae subsp. *americana* LMG 27715
		Brenneria salicis ATCC 15712
	Dickeya	*Dickeya aquatica* 174/2
		Dickeya chrysanthemi ATCC 11663
		Dickeya dadantii 3937
		Dickeya dadantii subsp. *dadantii* NCPPB 898
		Dickeya dadantii subsp. *dieffenbachiae* NCPPB 2976
		Dickeya dianthicola NCPPB 453
		Dickeya fangzhongdai DSM 101947
		Dickeya paradisiaca ATCC 33242
		Dickeya solani IPO 2222
		Dickeya zeae NCPPB 2538
	Lonsdalea	*Lonsdalea britannica* LMG 26267
		Lonsdalea iberica LMG 26264
		Lonsdalea populi LMG 27349
		Lonsdalea quercina ATCC 29281
	Pectobacterium	*Pectobacterium atrosepticum* ATCC 33260
		Pectobacterium betavasculorum ATCC 43762
		Pectobacterium carotovorum subsp. *carotovorum* ATCC 15713
		Pectobacterium parmentieri RNS 08.42.1A
		Pectobacterium polaris NIBIO 1006
		Pectobacterium wasabiae ATCC 43316
	Sodalis	*Sodalis praecaptivus* ATCC BAA-2554
Enterobacteriaceae	*Escherichia*	*Escherichia coli* K-12 substr. MG1655
	Salmonella	*Salmonella* Typhimurium LT2
Hafniaceae	*Hafnia*	*Hafnia alvei* ATCC 13337
Morganellaceae	*Morganella*	*Morganella morganii* subsp. *morganii* ATCC 25830
Yersiniaceae	*Yersinia*	*Yersinia enterocolitica* subsp. *enterocolitica* ATCC 9610

Table 3.2: Conservation of FliA binding site in *D. dadantii* 3937 compared to *E. coli* K-12 substr. MG1655. The consensus FliA binding site consists of TAAAGTTT-N11-GCCGATAA (15).

Gene/Operon[1]	**ASAP Feature ID[2]**	**Putative FliA binding site in *E. coli* (top) and *D. dadantii* (bottom)[4]**
fliC	ABE-0006402	**TAAAG**G**TT**gttttacgaca**G**A**CGATAA**C
	ABF-0019735	**TAAAG**G**TT**gttttacaccg**GCCGATA**CA
fliDST	ABE-0006407	**TAAA**C**TTT**gcgcaattcaga**CCGATAA**C
	ABF-0017963	**TAAA**C**TTT**tcgtatacagt**GCCGATAA**G
flgKL	ABE-0003660	**T**C**AAGT**CCggcgggtcgct**GCCGATAA**T
	ABF-0017941	**T**C**AAGTTT**tcagcccgcct**GCCGATA**GC
flgMN	ABE-0003634	**TAAAG**A**TT**acccgtcccttt**GCCGATAA**A
	ABF-0017928	**TAAAGT**AAtgctcaatgct**GCCGATA**TG
motAB-cheAW	ABE-0006299	**TAAA**C**TTT**cccagaatcct**GCCGATA**TT
	ABF-0018761	**TAAA**C**TTT**cgccgagcgtt**GCCGATA**TC
flxA	ABE-0005234	**TAAAG**A**TT**ttttgtgcat**GCCGATA**GT
	ABF-0014838[3]	**TAAAG**A**T**Accgcggcgctat**GCCGATAA**C
tolA	–	–
	ABF-0017664	**TAAAG**GCGgcgggcgcggt**GCCG**T**TA**TG
yjfB	–	–
	ABF-0018207	**TAAAGTTT**gaacagaaaagA**CCGATA**CC
–	–	–
	ABF-0016874	**TAAAG**A**TT**tggtcggagga**G**T**CGATA**TT
–	–	–
	ABF-0018601[3]	**TAAAG**A**TT**tattttacacttA**CCGATAA**C

[1]Gene(s) adjacent to FliA binding site.
[2]Gene ID is based on the single gene adjacent to FliA binding site.
[3]The motif contains a 12-nucleotide spacer.
[4]Bolded nucleotide corresponds to consensus FliA motif.

Table 3.3: Putative FliA motif in the promoters of 12 methyl-accepting chemotaxis protein genes in *D. dadantii* 3937 in comparison to *E. coli* K-12 substr. MG1655. The consensus FliA motif consists of TAAAGTTT-N11-GCCGATAA (15).

ASAP Feature ID	**Putative FliA Motif[2]**
***E. coli* K-12 substr. MG1655**	
ABE-0004744 (*trg*)	**TAAGT**AA**T**taccgtcaagt**GCCGAT**G**A**C
ABE-0006290 (*tar*) (part of the operon *tar-tap-cheRBYZ*)	**TAAAGTTT**ccccctcctt**GCCGATAA**C
ABE-0010087 (*aer*)	**TAAAG**A**T**Aaccgcagcggg**GCCGA**C**A**TA
ABE-0014282 (*tsr*)	**TAAAGTTT**ttcctttccag**GCCGA**A**AA**T
***D. dadantii* 3937**	
ABF-0014722[1]	**TAAAGTT**Cccgcttttattt**GCCGAT**GGC
ABF-0019050[1]	**TAAA**TC**T**CtctactgcagacA**CCGATAA**C
ABF-0018754 (upstream of *cheRBYZ*)	**TAA**TAAACcagccatattg**G**G**C**C**A**CTTC
ABF-0014824	**TAAAG**AACaacggcaggatT**CCGATAA**A
ABF-0016115	**TAAAGTTT**tactatggtgc**GCCGATA**GA
ABF-0016380	**TAAAG**AAGcccctatccag**G**T**CGATA**TC
ABF-0017896	**TAAA**TAAAaagacgaaatg**GCCGATA**TA
ABF-0018541	ATTT**GT**AAactgaagtatg**GCCGATA**TA
ABF-0018585	**TAA**TCCACcgcgttaacat**GCCGATAA**C
ABF-0018765	**T**G**AA**TG**TT**tcacattatat**GCCGAT**GTT
ABF-0019306	**TAAAG**GCAgtgggtcgttg**G**T**CGAT**T**A**A
ABF-0046680	**TAAAGTTT**ccgtgtcgacg**G**A**CGAT**G**A**A

[1]The motif contains a 12-nucleotide spacer.
[2]Bolded nucleotide corresponds to consensus FliA motif.

Table 3.4: Differentially expressed genes in *fliA* mutant compared to WT *D. dadantii* 3937 in MM-glu and MM-gly.

ASAP feature ID	**Gene name**	**Product**	**Log2 fold change in MM-glu[1]**	**Log2 fold change in MM-gly[1]**
ABF-0019721	*fliZ*	flagellar regulatory protein FliZ	1.38	1.09
ABF-0017930	*flgA*	flagella basal body P-ring formation protein FlgA	1.05	
ABF-0017931	*flgB*	flagellar basal body rod protein FlgB	1.65	
ABF-0017958	*fliE*	flagellar hook-basal body complex protein FliE	1.74	
ABF-0017957	*fliF*	flagellar M-ring protein FliF	1.16	
ABF-0047147	*gpV*	baseplate assembly protein; Type VI secretion system	1.29	
ABF-0016857	*traF*	conjugal transfer protein TraF	1.03	1.08
ABF-0019735	*fliC*	flagellin FliC; flagellar filament subunit	-5.81	-5.79
ABF-0017963	*fliD*	flagellar filament capping protein FliD	-1.74	-1.39
ABF-0017961	*fliS*	FliC chaperone protein FliS	-1.93	-1.48
ABF-0017960	*fliT*	FliD chaperone protein FliT	-1.75	-1.20
ABF-0017941	*flgK*	flagellar first hook-filament junction protein FlgK	-2.34	-1.14
ABF-0017942	*flgL*	flagellar hook-associated protein 3, FlgL	-1.74	
ABF-0017928	*flgM*	anti-sigma-28 factor FlgM	-2.82	-2.49
ABF-0047190	*flgN*	FlgK and FlgL chaperone protein FlgN	-1.87	-1.35
ABF-0014838	*flxA*	putative exported protein	-4.87	-4.22
ABF-0017665		methyl-accepting chemotaxis protein	-1.49	
ABF-0017097		methyl-accepting chemotaxis protein	-1.04	
ABF-0020167		methyl-accepting chemotaxis protein		-1.45
ABF-0019050		methyl-accepting chemotaxis protein		-1.16
ABF-0016874		hypothetical protein	-1.93	-1.02
ABF-0018207	*yjfB*	putative motility protein		-1.29
ABF-0018601		hypothetical protein	-1.87	
ABF-0017664	*tolA*	protein of unknown function	-2.66	-1.53

[1]False Discovery Rate (FDR) adjusted *p*-values are all $\leq$0.05.

Table 3.5: Relative expression of chemotaxis related genes in WT *D. dadantii* 3937 in MM-glu compared to MM-gly.

ASAP feature ID	**Gene name**	**Product**	**Average normalized count values in MM-glu**	**log2 fold change**	**FDR-adjusted *p*-value**
ABF-0018757	*cheA*	histidine kinase CheA	7402.50	-0.19	0.60
ABF-0019072	*cheV*	adaptor protein CheV	2512.16	-0.10	0.86
ABF-0018749	*cheZ*	phosphatase CheZ	3708.64	-0.54	0.06
ABF-0014726	*aerA*	aerotaxis	179.61	-0.34	0.21
ABF-0019309	*aerC*	aerotaxis	684.64	-0.12	0.65
ABF-0014618		methyl-accepting chemotaxis protein	161.41	-1.19	0.00
ABF-0020167		methyl-accepting chemotaxis protein	221.15	-0.86	0.00
ABF-0018541		methyl-accepting chemotaxis protein	1018.42	-0.70	0.19
ABF-0019306		methyl-accepting chemotaxis protein	38.87	-0.57	0.09
ABF-0018511		methyl-accepting chemotaxis protein	293.82	-0.46	0.01
ABF-0018765		methyl-accepting chemotaxis protein	710.62	-0.46	0.34
ABF-0019050		methyl-accepting chemotaxis protein	197.08	-0.46	0.24
ABF-0018502		methyl-accepting chemotaxis protein	626.06	-0.43	0.00
ABF-0020431		methyl-accepting chemotaxis protein	1504.62	-0.15	0.84
ABF-0017419		methyl-accepting chemotaxis protein	65.64	-0.14	0.75
ABF-0017090		methyl-accepting chemotaxis protein	156.31	-0.13	0.80
ABF-0018751	*cheB*	methylesterase CheB	2100.08	0.01	0.99
ABF-0018752	*cheR*	methyltransferase CheR	5342.07	0.72	0.21
ABF-0018755	*cheW*	scaffolding protein CheW	8014.85	0.02	0.97
ABF-0014843	*aerB*	aerotaxis	15.97	0.14	0.83
ABF-0017674		methyl-accepting chemotaxis protein	280.52	1.61	0.00
ABF-0017537		methyl-accepting chemotaxis protein	378.53	1.43	0.00
ABF-0017665		methyl-accepting chemotaxis protein	56.25	1.34	0.00
ABF-0015600		methyl-accepting chemotaxis protein	115.12	1.03	0.00
ABF-0016380		methyl-accepting chemotaxis protein	84.73	1.01	0.00
ABF-0014536		methyl-accepting chemotaxis protein	575.37	0.95	0.18
ABF-0019718		methyl-accepting chemotaxis protein	226.22	0.88	0.07

ABF-0016436		methyl-accepting chemotaxis protein	99.26	0.68	0.17
ABF-0017896		methyl-accepting chemotaxis protein	405.32	0.62	0.41
ABF-0016115		methyl-accepting chemotaxis protein	954.88	0.59	0.14
ABF-0019790		methyl-accepting chemotaxis protein	583.29	0.53	0.11
ABF-0014722		methyl-accepting chemotaxis protein	33.38	0.51	0.31
ABF-0014824		methyl-accepting chemotaxis protein	49.45	0.48	0.14
ABF-0017668		methyl-accepting chemotaxis protein	154.78	0.39	0.55
ABF-0019851		methyl-accepting chemotaxis protein	904.13	0.38	0.60
ABF-0020252		methyl-accepting chemotaxis protein	211.43	0.36	0.45
ABF-0018754		methyl-accepting chemotaxis protein	11257.85	0.33	0.56
ABF-0017824		methyl-accepting chemotaxis protein	20.57	0.31	0.53
ABF-0017097		methyl-accepting chemotaxis protein	492.45	0.31	0.51
ABF-0019855		methyl-accepting chemotaxis protein	213.03	0.31	0.62
ABF-0017672		methyl-accepting chemotaxis protein	57.06	0.29	0.46
ABF-0017662		methyl-accepting chemotaxis protein	144.47	0.28	0.52
ABF-0015513		methyl-accepting chemotaxis protein	280.64	0.27	0.17
ABF-0018892		methyl-accepting chemotaxis protein	244.00	0.23	0.63
ABF-0017863		methyl-accepting chemotaxis protein	185.41	0.21	0.56
ABF-0015603		methyl-accepting chemotaxis protein	121.03	0.19	0.75
ABF-0019852		methyl-accepting chemotaxis protein	459.17	0.16	0.80
ABF-0016585		methyl-accepting chemotaxis protein	107.83	0.14	0.70
ABF-0019858		methyl-accepting chemotaxis protein	148.79	0.11	0.86
ABF-0046680		methyl-accepting chemotaxis protein	270.18	0.11	0.72
ABF-0016979		methyl-accepting chemotaxis protein	1944.53	0.08	0.90
ABF-0015168		methyl-accepting chemotaxis protein	169.86	0.07	0.80
ABF-0018585		methyl-accepting chemotaxis protein	352.02	0.02	0.98

Table 3.6: Conservation of sigma factors in Pectobacteriaceae family compared to closely related animal pathogen *Escherichia coli.*

Sigma Factor	**RpoD**	**RpoE**	**RpoH**	**RpoN**	**RpoS**	**FliA**	**HrpL**	**FecI**
E. coli ATCC 11775	+	+	+	+	+	+	-	+
Dickeya dadantii 3937	+	+	+	+	+	+	+	+
Dickeya aquatica 174/2	+	+	+	+	+	+	+	+
Dickeya chrysanthemi ATCC 11663	+	+	+	+	+	+	+	-
Dickeya dadantii subsp. *dadantii* NCPPB 898	+	+	+	+	+	+	+	+
Dickeya dadantii subsp. *dieffenbachiae* NCPPB 2976	+	+	+	+	+	+	+	+
Dickeya dianthicola NCPPB 453	+	+	+	+	+	+	+	-
Dickeya fangzhongdai DSM 101947	+	+	+	+	+	+	+	+
Dickeya paradisiaca ATCC 33242	-	+	+	+	+	+	-	+
Dickeya solani IPO 2222	+	+	+	+	+	+	+	+
Dickeya zeae NCPPB 2538	+	+	+	+	+	+	+	-
Pectobacterium atrosepticum ATCC 33260	+	+	+	+	-	+	+	+
Pectobacterium betavasculorum ATCC 43762	+	+	+	+	+	+	+	+
Pectobacterium brasiliensis 1692	+	+	+	+	+	+	+	+
Pectobacterium carotovorum subsp. *carotovorum* ATCC 15713	+	+	+	+	+	+	+	+
Pectobacterium carotovorum subsp. *odoriferum* NCPPB 3839	+	+	+	+	+	+	+	+
Pectobacterium parmentieri RNS 08.42.1A	+	+	+	+	+	+	+	+
Pectobacterium polaris NIBIO 1006	+	+	+	+	+	+	+	+
Pectobacterium wasabiae ATCC 43316	+	+	+	+	+	+	+	+

‘+’ indicates the presence of the sigma factor.
‘-‘ represents the absence of the sigma factor.

Table 3.7: A subset of 84 out of 224 genes differentially up-regulated in WT *D. dadantii* 3937 in MM-glu vs MM-gly.

ASAP feature ID	Gene name	Product	log2 Fold Change[1]
	Plant-microbe interactions		
ABF-0018072	*budA*	alpha-acetolactate decarboxylase	7.07
ABF-0018071	*budB*	acetolactate synthase	4.06
ABF-0019339	*budC*	acetoin reductase	2.86
ABF-0018073	*budR*	transcriptional regulator	2.03
ABF-0018252	*cbsA*	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	2.96
ABF-0018254	*cbsB*	isochorismatase	3.69
ABF-0018256	*cbsC/entC*	isochorismate synthase EntC	3.79
ABF-0018255	*cbsE*	2,3-dihydroxybenzoate-AMP ligase; chrysobactin biosynthesis	3.90
ABF-0019151	*cbsF*	chrysobactin synthetase	2.38
ABF-0019153	*cbsH*	enterochelin esterase; chrysobactin oligopeptidase	4.43
ABF-0018251	*cbsP*	acyl-CoA esterase; positive regulation by HrpL; expression of this gene was decreased in an hrpL null mutant as determined by microarray analysis.	2.17
ABF-0019156	*fct*	ferrichrysobactin receptor	2.74
ABF-0016754	*fusA*	TonB-dependent receptor - iron acquisition from plant ferredoxin	2.63
ABF-0020373	*prtA*	serine 3-dehydrogenase	2.50
ABF-0047107	*prtB*	serine 3-dehydrogenase	2.06
ABF-0020371	*prtC*	serine 3-dehydrogenase; serralysin-like metalloprotease	3.74
ABF-0019593	*hrpA*	HrpA family pilus protein T3SS	3.66
ABF-0020784	*hrpN*	Hairpin hrpN T3SS	3.03
ABF-0019143	*avrL*	avirulence protein; AvrL is secreted by the Out system; promoter is bound by pecS	3.20
ABF-0015381	*avrM*	avirulence protein	3.55
ABF-0018556	*wcaJ/cpsA*	undecaprenyl-phosphate glucose phosphotransferase; capsular polysaccharide synthesis enzyme CpsA	3.40
ABF-0018555		capsular polysaccharide biosynthesis protein	3.32
ABF-0018554		capsular polysaccharide biosynthesis protein	3.06

ABF-0018635		capsular polysaccharide biosynthesis glycosyl transferase	2.90
ABF-0018553		capsular polysaccharide biosynthesis protein	2.85
ABF-0018636		alpha-D-GlcNAc alpha-1,2-L-rhamnosyltransferase (capsular polysaccharide biosynthesis glycosyl transferase)	2.80
ABF-0018552		capsular polysaccharide biosynthesis protein	2.76
ABF-0018634		capsular polysaccharide biosynthesis glycosyl transferase	2.67
ABF-0018639	*wcaB*	serine acetyltransferase; colanic acid biosynthesis acetyltransferase; putative capsular polysacharide biosynthesis transferase	1.71
ABF-0014642	*yoaJ*	Expansin-YoaJ; (family of closely related nonenzymatic proteins found in plant cell wall); Endoglucanase	3.28
ABF-0019026	*xynA*	xylanase	1.96
ABF-0019503	*ybjX*	VirK protein (resistance to antimicrobial peptides)	1.24
	Microbe-microbe interactions		
ABF-0017270	*hcpA*	haemolysin-Coregulated Protein (Hcp) T6SS	2.40
ABF-0018743	*hcpB*	haemolysin-Coregulated Protein (Hcp) T6SS	3.32
ABF-0015593	*hcp*	haemolysin-Coregulated Protein (Hcp) T6SS	1.81
ABF-0017870		haemolysin-Coregulated Protein (Hcp) T6SS	2.45
ABF-0017269	$vgrG_A$	valine-glycine rich (Vgr) repeat proteins T6SS	2.18
ABF-0018744	$vgrG_B$	valine-glycine rich (Vgr) repeat proteins T6SS	2.98
ABF-0016665	*cytA*	delta-endotoxin CytA	2.67
ABF-0016662	*cytD*	type-1Aa cytolytic delta-endotoxin	1.05
ABF-0016053	*yafN*	Type II toxin-antitoxin system, antitoxin Phd/YefM	1.32
ABF-0016054	*yafO*	Toxin YafO, type II toxin-antitoxin system	1.20
	Motility-related proteins		
ABF-0019839	*rhlA*	3-hydroxydecanoyl-ACP:CoA transacylase (catalyzes the production of surfactant for swarming motility)	5.08
ABF-0018581	*flhE*	flagellar protein FlhE	3.63
ABF-0017674		methyl-accepting chemotaxis protein	1.61
ABF-0017537		methyl-accepting chemotaxis protein	1.43
ABF-0017665		methyl-accepting chemotaxis protein	1.34

ABF-0015600		methyl-accepting chemotaxis protein	1.03
ABF-0016380		methyl-accepting chemotaxis protein	1.01
	ABC transporters		
ABF-0017010	*yhdW*	amino acid ABC transporter substrate-binding protein	3.78
ABF-0017008	*yhdX*	amino acid ABC transporter permease	1.27
ABF-0017007	*yhdY*	amino acid ABC transporter permease	1.82
ABF-0017006	*yhdZ*	amino acid ABC transporter ATP-binding protein	2.20
ABF-0019444		dipeptide ABC transporter substrate-binding protein	3.71
ABF-0019448		dipeptide ABC transporter ATP-binding protein	2.74
ABF-0019447		dipeptide ABC transporter ATP-binding protein	2.70
ABF-0019446		dipeptide ABC transporter permease	2.62
ABF-0019445		dipeptide ABC transporter permease	2.42
ABF-0017399	*nasD*	nitrate ABC transporter ATP-binding protein	3.47
ABF-0017398	*nasE*	nitrate ABC transporter permease protein	2.68
ABF-0017397	*nasF*	nitrate ABC transporter substrate-binding protein	2.86
ABF-0020780		polyamine ABC transporter ATP-binding protein	2.92
ABF-0019509	*hmuU*	iron ABC transporter	2.42
ABF-0015909	*hmuV*	heme ABC transporter ATP-binding protein	2.58
ABF-0019572		peptide uptake ABC transporter ATP-binding protein	2.54
ABF-0019570		peptide uptake ABC transporter ATP-binding protein	2.10
ABF-0016756	*fusD*	iron-sulphur ABC transporter ATP-binding protein	2.44
ABF-0015620	*potB*	spermidine/putrescine ABC transporter permease	2.08
ABF-0015621	*potC*	spermidine/putrescine ABC transporter substrate-binding protein	2.00
ABF-0015623	*potD*	spermidine/putrescine ABC transporter substrate-binding protein	2.25
ABF-0016875	*macB*	macrolide exporter ABC transporter permease	2.20
ABF-0019148	*cbuC*	ferric chrysobactin uptake ABC transporter ATP-binding subunit	1.74
ABF-0019146	*cbuD*	ferric chrysobactin uptake ABC transporter permease	2.06
ABF-0016441		putative ABC transporter	2.05
ABF-0019222		zinc uptake ABC transporter ATP-binding protein	2.05
ABF-0019223		zinc uptake ABC transporter permease	2.03

ABF-0017107		maltose/maltodextrin uptake ABC transporter permease	1.94
ABF-0018788		amino acid ABC transporter permease	1.90
ABF-0016124		urea ABC transporter substrate-binding protein	1.78
ABF-0017106		maltose/maltodextrin uptake ABC transporter ATP-binding protein	1.68
ABF-0017108		maltose/maltodextrin uptake ABC transporter permease	1.52
ABF-0014704	*hisQ*	histidine/lysine/arginine/ornithine ABC transporter permease	1.50
ABF-0014703	*hisJ*	histidine/lysine/arginine/ornithine ABC transporter substrate-binding protein	1.19
ABF-0020873	*hisM*	histidine/lysine/arginine/ornithine ABC transporter permease	1.19

[1]False Discovery Rate (FDR) adjusted *p*-values are all $\leq$0.05.

Table 3.8: A subset of 73 out of 103 genes differentially down-regulated in WT *D. dadantii* 3937 in MM-glu vs MM-gly.

ASAP feature ID	Gene name	Product	log2 Fold Change[1]
	Plant-microbe interactions		
ABF-0019587	*hrpE*	T3SS HrpE	-1.25
ABF-0016066	*vfmX*	putative auto-transporter adhesin; VfmX	-1.18
	Motility-related proteins		
ABF-0017931	*flgB*	flagellar basal body rod protein FlgB	-1.44
ABF-0046587	*flgC*	flagellar basal body rod protein FlgC	-1.58
ABF-0017934	*flgD*	flagellar basal body rod modification protein FlgD	-1.57
ABF-0017935	*flgE*	flagellar hook protein FlgE	-1.61
ABF-0017936	*flgF*	flagellar basal body rod protein FlgF	-1.42
ABF-0017937	*flgG*	flagellar basal body rod protein FlgG	-1.34
ABF-0017938	*flgH*	flagellar basal body L-ring protein; flgH	-1.26
ABF-0017958	*fliE*	flagellar hook-basal body complex protein FliE	-1.21
ABF-0017957	*fliF*	flagellar M-ring protein FliF	-1.09
ABF-0014618		methyl-accepting chemotaxis protein	-1.19
	Metabolism-related proteins		
ABF-0015033	*glpA*	anaerobic sn-glycerol-3-phosphate dehydrogenase subunit A; GlpA	-1.78
ABF-0015034	*glpB*	anaerobic glycerol-3-phosphate dehydrogenase subunit B; GlpB	-2.16
ABF-0015037	*glpC*	anaerobic sn-glycerol-3-phosphate dehydrogenase subunit C; GlpC	-2.12
ABF-0015062	*glpD*	aerobic glycerol-3-phosphate dehydrogenase; GlpD	-4.27
ABF-0020399	*glpF*	aquaporin; glycerol uptake facilitator protein; GlpF	-3.94
ABF-0020398	*glpK*	glycerol kinase; GlpK	-4.75
ABF-0015031	*glpT*	MFS transporter; glycerol-3-phosphate transporter; GlpT	-4.36
ABF-0015029	*glpQ*	glycerophosphoryl diester phosphodiesterase; GlpQ	-2.55
ABF-0019893	*ppsA*	phosphoenolpyruvate synthase	-2.80
ABF-0018196	*ganA*	galactosidase; periplasmic 1,4-beta-endogalactanase	-1.81
ABF-0018198	*ganB*	beta-galactosidase; 1,4-beta-exogalactanase	-2.15
ABF-0018898	*rhaD*	rhamnulose-1-phosphate aldolase	-2.50

ABF-0016431	*iolA*	malonate semialdehyde dehydrogenase	-1.37
ABF-0016433	*iolD*	thiamine pyrophosphate enzyme	-1.62
ABF-0016453	*iolC*	5-dehydro-2-deoxygluconokinase	-1.49
ABF-0016452	*iolG1*	myo-inositol 2-dehydrogenase	-1.42
ABF-0018066	*gcvH*	glycine cleavage system protein H	-1.66
ABF-0018065	*gcvP*	glycine dehydrogenase	-1.70
ABF-0020198	*sfcA/maeA*	NAD-dependent malic enzyme	-1.24
ABF-0014791	*atpF*	F0F1 ATP synthase subunit B (oxidative stress response protein)	-1.07
ABF-0017092	*sodC*	superoxide dismutase	-1.26
	Regulation		
ABF-0018901	*rhaS*	transcriptional activator RhaS	-1.75
ABF-0018902	*rhaR*	AraC family transcriptional regulator; RhaR	-2.01
ABF-0016098	*xylR*	XylR family transcriptional regulator; regulates xylose metabolism genes	-1.66
ABF-0016089	*pecS*	PecS (MarR family transcriptional regulator)	-1.07
ABF-0020446	*rpoS*	RNA polymerase sigma factor RpoS	-1.04
	Transport		
ABF-0019925	*uhpT*	hexose phosphate transporter; UhpT	-4.68
ABF-0018415	*dppA*	dipeptide ABC transporter substrate-binding protein	-2.04
ABF-0018416	*dppB*	dipeptide ABC transporter permease	-1.58
ABF-0018417	*dppC*	dipeptide ABC transporter permease	-1.70
ABF-0018418	*dppD*	dipeptide ABC transporter ATP-binding protein	-1.56
ABF-0017600	*dppF*	dipeptide ABC transporter ATP-binding protein	-1.88
ABF-0018199	*ganC*	PTS sugar transporter	-1.77
ABF-0018192	*ganE/malE*	arabinogalactan uptake ABC transporter perisplasmic binding protein	-1.77
ABF-0018193	*ganF/malF*	arabinogalactan ABC transporter permease	-2.11
ABF-0018195	*ganG/malG*	arabinogalactan ABC transporter permease	-2.22
ABF-0018200	*ganL*	maltoporin	-1.77
ABF-0016469	*mltF*	ABC transporter substrate-binding protein	-1.78
ABF-0016450		sugar ABC transporter permease	-1.42
ABF-0016448		sugar ABC transporter ATP-binding protein	-1.37

ABF-0020639	*oppA*	oligopeptide ABC transporter substrate-binding protein	-1.23
ABF-0020638	*oppB*	oligopeptide ABC transporter permease	-1.07
ABF-0020637	*oppC*	oligopeptide ABC transporter permease	-1.30
ABF-0020636	*oppD*	oligopeptide ABC transporter ATP-binding protein	-1.25
ABF-0020635	*oppF*	oligopeptide ABC transporter ATP-binding protein	-1.21
ABF-0019350	*ytfQ*	sugar ABC transporter substrate-binding protein	-1.20
ABF-0017659	*livG*	branched-chain amino acid ABC transporter ATP-binding protein	-1.15
ABF-0017657	*livH*	branched-chain amino acid ABC transporter permease	-1.16
ABF-0017656	*livK*	branched chain amino acid ABC transporter substrate-binding protein	-1.17
ABF-0017658	*livM*	branched-chain amino acid ABC transporter permease	-1.14
ABF-0015566	*araF*	arabinose ABC transporter substrate-binding protein	-1.03
ABF-0047129	*togB*	peptic oligomers ABC transporter substrate-binding protein	-1.02
ABF-0018137	*rbsA*	ribose ABC transporter ATP-binding protein	-1.00
ABF-0017796	*tctA*	tricarboxylic acid transport membrane protein	-1.76
ABF-0017795	*tctB*	tricarboxylic acid transport membrane protein	-2.58
ABF-0017794	*tctC*	tricarboxylic acid transport membrane protein	-2.18
ABF-0019695	*ansP*	L-asparagine transporter	-1.49
ABF-0020858	*aroP*	aromatic amino acid transporter AroP	-1.19
ABF-0019249		carbon starvation protein	-1.22
	Hypothetical proteins		
ABF-0018593	*yjbJ*	hypothetical protein	-2.25
ABF-0015573		hypothetical protein	-1.13

[1]False Discovery Rate (FDR) adjusted *p*-values are all $\leq$0.05.

Table 3.9: ABC transporter genes that were differentially regulated* in MM-glu compared to MM-gly.

	ABC Family	Transporter operon	Gene	Accession Number	Function	Log_2Fold change
Up-regulated	2	ABF-0019444 to ABF-0019448	-	ABF-0019444	dipeptide uptake	3.71
			-	ABF-0019445		2.42
			-	ABF-0019447		2.70
			-	ABF-0019448		2.74
		ABF-0019568 to ABF-0019572	-	ABF-0019570	peptide uptake	2.10
			-	ABF-0019572		2.54
	3	ABF-0016441	-	ABF-0016441	putative transporter	2.05
	4	HisJMPQ	*hisJ*	ABF-0014703	lysine/arginine/ornithine uptake	1.19
			hisM	ABF-0020873		1.19
			hisQ	ABF-0014704		1.50
		YhdWXYZ	*yhdZ*	ABF-0017006	putative amino acid uptake	2.20
			yhdY	ABF-0017007		1.82
			yhdX	ABF-0017008		1.27
			yhdW	ABF-0017010		3.78
		ABF-0018788 to ABF-0018791	-	ABF-0018788	amino acid uptake	1.90
	5	PotABCD	*potB*	ABF-0015620	spermidine/putrescine uptake	2.08
			potC	ABF-0015621		2.00
			potD	ABF-0015623		2.25
		ABF-0017106 to ABF-0017109	-	ABF-0017106	maltose/maltodextrin uptake	1.68
			-	ABF-0017107		1.94
			-	ABF-0017108		1.52
		NasDEF	*nasF*	ABF-0017397	nitrite uptake	2.86

			nasE	ABF-0017398		2.68
			nasD	ABF-0017399		3.47
		ABF-0020780	-	ABF-0020780	polyamine uptake	1.19
	6	FusCD	*fusD*	ABF-0016756	iron-sulphur uptake	2.44
	8	HmuSTUV	*hmuV*	ABF-0015909	heme uptake	2.42
			hmuU	ABF-0019509		2.58
		CbuBCDG	*cbuD*	ABF-0019146	ferric chrysobactin uptake	2.06
			cbuC	ABF-0019148		1.74
		ABF-0019222, ABF-0019223, and ABF-0046525	-	ABF-0019222	zinc uptake	2.05
			-	ABF-0019223		2.03
	9	MacAB	*macB*	ABF-0016875	macrolide exporter	2.20
Down-regulated	1	AraFGH	*araH*	ABF-0015566	L-arabinose, fructose, xylose uptake	-1.03
		ABF-0016447 to ABF-0016450	-	ABF-0016450	sugar uptake	-1.42
		RbsABC	*rbsA*	ABF-0018137	ribose uptake	-1.00
	2	DppABCDF	*dppA*	ABF-0018415	dipeptide uptake	-2.04
			dppB	ABF-0018416		-1.58
			dppC	ABF-0018417		-1.70
			dppD	ABF-0018418		-1.56
			dppF	ABF-0017600		-1.88
		OppABCDF	*oppF*	ABF-0020635	oligopeptide uptake	-1.21
			oppD	ABF-0020636		-1.25
			oppC	ABF-0020637		-1.30
			oppB	ABF-0020638		-1.07
			oppA	ABF-0020639		-1.23
	5	GanEFGK	*ganE*	ABF-0018192	galactan uptake	-1.77
			ganF	ABF-0018193		-2.11
			ganG	ABF-0018195		-2.22

		YtfQRT	*ytfQ*	ABF-0019350	galactofuranose uptake	-1.20
		TogMNAB	*togB*	ABF-0047129	pectic oligomers	-1.02
	10	LivFGHKM	*livK*	ABF-0017656	leucine, isoleucine, and valine uptake	-1.17
			livH	ABF-0017657		-1.16
			livM	ABF-0017658		-1.14
			livG	ABF-0017659		-1.15

*: FDR adjusted *P*-values are all <0.01.

Table 3.10: Identification of putative CRP binding site in the upstream region of 21 genes/operons in *D. dadantii* 3937 in comparison to *E. coli* K-12 substr. MG1655. The typical CRP binding site consists of TGTGA-N(6-8)-TCACA (70).

Gene/Operon[1]	**ASAP Feature ID[2]**	**Putative CRP binding site in *E. coli*[3] (top) and *D. dadantii* (bottom)**
glpABC	ABE-0007422	AATGAC**G**Catgaaa**TCAC**GTTT
	ABF-0015033	CTGA**GTGA**tttgag**TCACA**TAT
glpD	ABE-0011191	TAA**TGT**T**A**tacata**TCAC**TCTA
	ABF-0015062	AAC**TGT**T**A**gctgca**TCAC**GCGT
glpFK	ABE-0012828	TTT**TA****TGA**cgaggcA**CACA**CAT
	ABF-0020399	TTTCT**TGA**cgtgtcA**CACA**CAT
glpTQ	ABE-0007401	ATG**TGTG**Cggcaat**TCACA**TTT
	ABF-0015031	ATA**TGTGA**ctcaaa**TCAC**TCAG
rhaSR	ABE-0012745	TGA**TGTGA**tgctcaC**C**G**CA**TTT TTTCC**TGA**aaattcA**C**G**C**TGTA
	ABF-0018901	GAC**TGTGA**atcagc**TCACA**GTT
gcvTHP	ABE-0009536[5]	**TGTGA**tgggta**TCA**A**A**
	ABF-0018067	CGTG**G**G**G**Cctggcg**TCACA**ACG
uhpT	ABE-0011981	AAGC**GTGA**tgcatc**TCAC**CTTT
	ABF-0019925	TAGC**GTGA**ttcgct**TCAC**CCTT
araFGH	ABE-0006330	CGA**TGTGA**tattgc**TC**T**C**CTAT
	ABF-0015566	GAA**TGTGA**taatgc**TC**T**C**TCTT
dppABCDF	ABE-0011576[5]	**TGT**CGtgtagg**TC**GA**A**
	ABF-0018415	TGTGA**TGA**caacaa**TC**G**C**GATG
malEFG/ganEFG	ABE-0013200	TTC**TGT**A**A**cagaga**TCACA**CAA
	ABF-0018192[4]	AAA**TGT**T**A**tacct**TCACA**TGA
oppABCDF	ABE-0004172	Unknown if regulated by CRP
	ABF-0020639	AGA**TGTGA**ttgagcGGC**CA**TGT
ytfQRT-yjfF	ABE-0013829[5]	AT**TG**Cgttaca**TCACA**

	ABF-0019350	ATATT**TG**Cgtggag**TCACA**GAT
ansP	ABE-0004845	Unknown if regulated by CRP
	ABF-0019695	GGC**TGTGA**cacaacAGC**CA**ATT
aroP	ABE-0000384	Unknown consensus but this operon is known to be regulated by CRP
	ABF-0020858	CGG**TG**ATCatgaca**TCACA**CTT
tctCBA	-	–
	ABF-0017794	TTAA**G**C**GA**cgccaa**TCACA**TCC
ppsA	ABE-0005678[5]	**TG**C**GA**ggtgtg**TCACA**
	ABF-0019893	CATG**G**C**G**caatatt**TCACA**CCC
iolAD	-	–
	ABF-0016431	AAA**T**C**T**ATcctgga**TCACA**ATA
sfcA/maeA	ABE-0004931	Unknown if regulated by CRP
	ABF-0020198	ATA**TGTGA**ctttccC**CACA**TAT
yjbJ	ABE-0013247	Unknown if regulated by CRP
	ABF-0018593	GTCATATTtttccg**TCACA**CTT
	-	–
	ABF-0015573	CGGGT**T**TGtgcggc**TCACA**GAT
cstA	ABE-0002061	CGGA**GTGA**tcgagt**T**A**ACA**TTG
	ABF-0019249	AAA**T**A**T**AGctttgt**TCACA**TTT

[1]Gene(s) adjacent to CRP binding site
[2]Gene ID is based on the single gene adjacent to CRP binding site.
[3]The CRP binding sites were obtained from (52, 63)
[4] The motif contains a 5-nucleotide spacer.
[5]The motif is based on Supplementary Table 3 from (52), however, we could not locate the motif upstream of the gene.

Table 3.11: Comparison of glycerol-regulated genes found in this study with PecS regulated genes (64).

ASAP feature ID	**Gene name**	**Product**	**Log2 Fold Change in MM-glu vs MM-gly in WT background[1]**	**Log2 Fold Change *pecS* mutant vs WT (6h post-inoculation in *Arabidopsis*)[1]**
	Plant-microbe interactions			
ABF-0015381	*avrM*	avirulence protein	3.55	11.75
ABF-0019143	*avrL*	avirulence protein	3.20	7.46
ABF-0020373	*prtA*	serine 3-dehydrogenase	2.50	11.10
ABF-0047107	*prtB*	serine 3-dehydrogenase	2.06	15.92
ABF-0020371	*prtC*	serine 3-dehydrogenase	3.74	2.85
ABF-0014642	*yoaJ*	expansin	3.28	2.92
ABF-0020784	*hrpN*	T3SS hairpin	3.03	5.12
ABF-0019339	*budC*	acetoin reductase	2.86	
ABF-0016754	*fusA*	TonB-dependent receptor involved in iron acquisition from plant ferredoxin	2.63	-3.19
ABF-0018556	*wcaJ/cpsA*	undecaprenyl-phosphate glucose phosphotransferase; capsular polysaccharide synthesis enzyme	3.40	
ABF-0018555		capsular polysaccharide biosynthesis protein	3.32	
ABF-0018554		capsular polysaccharide biosynthesis protein	3.06	
ABF-0018635		capsular polysaccharide biosynthesis glycosyl transferase	2.90	
ABF-0018552		capsular polysaccharide biosynthesis protein	2.76	
ABF-0018634		capsular polysaccharide biosynthesis glycosyl transferase	2.67	
	Motility-related proteins			
ABF-0019839	*rhlA*	3-hydroxydecanoyl-ACP:CoA transacylase (catalyzes surfactant production for swarming motility)	5.08	10.16

ABF-0018581	*flhE*	flagellar swarming protein FlhE	3.63	4.64
ABF-0017674		methyl-accepting chemotaxis protein	1.61	5.15
ABF-0015600		methyl-accepting chemotaxis protein	1.03	2.54
ABF-0017936	*flgF*	flagellar basal body rod protein FlgF	-1.42	2.34
	Secretion systems			
ABF-0018743	*hcpB*	haemolysin-coregulated protein T6SS	3.32	7.42
ABF-0017270	*hcpA*	haemolysin-coregulated protein T6SS	2.40	4.85
ABF-0015593	*hcp*	haemolysin-coregulated protein T6SS	1.81	3.29
ABF-0017269	$vgrG_A$	valine-glycine-repeat capping protein T6SS	2.18	2.96
	Regulation			
ABF-0020518	*nac*	LysR family transcriptional regulator; nitrogen assimilation control; represses gdhA, serA, asnC and asnA transcription	4.26	
ABF-0018459		bacteriophage transcription regulator protein	1.08	-3.54
ABF-0016089	*pecS*	PecS	-1.07	9.05
ABF-0020446	*rpoS*	RNA polymerase stationary sigma factor RpoS	-1.04	-2.32
	Transport			
ABF-0017399	*nasD*	nitrate ABC transporter ATP-binding domain	3.47	3.12
ABF-0017397	*nasF*	nitrate ABC transporter substrate-binding domain	2.86	2.74
ABF-0015623	*potD*	spermidine/putrescine ABC transporter substrate-binding protein	2.25	3.03
ABF-0015620	*potB*	spermidine/putrescine ABC transporter permease	2.08	3.06
ABF-0015621	*potC*	spermidine/putrescine ABC transporter substrate-binding protein	2.00	3.40
ABF-0017010	*yhdW*	amino acid ABC transporter substrate-binding protein	3.78	
ABF-0017006	*yhdZ*	arginine ABC transporter ATP-binding protein	2.20	
ABF-0017007	*yhdY*	amino acid ABC transporter permease	1.82	

ABF-0019444		peptide ABC transporter substrate-binding protein	3.71	
ABF-0016441		putative ABC transporter	2.05	-3.86
ABF-0016124		substrate-binding component of an ABC superfamily urea utilization-associated transporter	1.78	2.28
ABF-0017108		maltose/maltodextrin uptake ABC transporter permease	1.52	-5.02
ABF-0016448		ATP-binding component of ABC superfamily sugar transporter	-1.37	
ABF-0017775		Fe2+ transport system protein A	2.11	-4.76
ABF-0017773	*feoA*	iron transporter	2.00	-2.01
ABF-0017803	*azlD*	branched amino acid transporter	1.78	-2.91
ABF-0019695	*ansP*	amino acid permease; L-asparagine transporter	-1.49	3.48
	Enzymes and metabolism-related proteins			
ABF-0019365	*yjiY/cstA*	carbon starvation protein A	1.49	4.27
ABF-0017401	*nirB*	nitrite reductase	1.53	7.34
ABF-0017402	*nasA*	nitrate reductase large subunit	1.04	4.29
ABF-0020585	*pta*	phosphate acetyltransferase	1.52	2.09
ABF-0019243	*aceE*	pyruvate dehydrogenase	1.49	2.11
ABF-0019241	*aceF*	pyruvate dehydrogenase complex dihydrolipoyllysine-residue acetyltransferase	1.15	4.20
ABF-0019960	*kdgA*	KHG/KDPG aldolase	1.12	3.52
ABF-0020586	*ackA*	acetate kinase and propionate kinase 2	1.07	2.08
ABF-0016308	*glnL/ntrB*	two-component system sensor histidine kinase GlnL/NtrB	1.01	2.36
ABF-0016233		4-hydroxyphenylpyruvate dioxygenase	1.67	
ABF-0020770		acyltransferase	1.67	
ABF-0020757	*ynfK*	dethiobiotin synthase	1.74	-3.38
ABF-0019893	*ppsA*	phosphoenolpyruvate synthase	-2.80	-4.57

ABF-0015037	*glpC*	anaerobic sn-glycerol-3-phosphate dehydrogenase subunit C	-2.12	-5.29
ABF-0015034	*glpB*	anaerobic sn-glycerol-3-phosphate dehydrogenase subunit B	-2.16	-17.98
ABF-0020765		glycosyl transferase	-1.27	-6.19
ABF-0016433	*iolD*	3D-(3,5/4)-trihydroxycyclohexane-1,2-dione acylhydrolase (decyclizing); Thiamine pyrophosphate enzyme	-1.62	
ABF-0016453	*iolC*	5-dehydro-2-deoxygluconokinase	-1.49	
ABF-0016452	*iolG1*	myo-inositol-2-dehydrogenase	-1.42	
	Hypothetical proteins			
ABF-0017268		hypothetical protein; protein DcrB	2.10	2.64
ABF-0020433		hypothetical protein	3.22	3.94
ABF-0018745		hypothetical protein	2.98	2.53
ABF-0020017		putative methoxymalonyl-ACP biosynthesis protein FkbH	2.55	2.03
ABF-0016874		hypothetical protein	2.28	2.12
ABF-0015852		hypothetical protein	1.08	2.85
ABF-0017176		DUF4879 domain-containing protein	3.43	
ABF-0015081		hypothetical protein	3.43	
ABF-0018551		hypothetical protein	2.77	
ABF-0020093		membrane protein	2.42	
ABF-0020769		hypothetical protein	1.99	
ABF-0015067		hypothetical protein	1.84	
ABF-0046541		hypothetical protein	1.55	
ABF-0015544		DUF4879 domain-containing protein	1.49	
ABF-0017026		putative DNA-binding protein	1.04	
ABF-0018593	*yjbJ*	hypothetical protein	-2.25	-21.90

[1]False Discovery Rate (FDR) adjusted *p*-values are all $\leq$0.05.

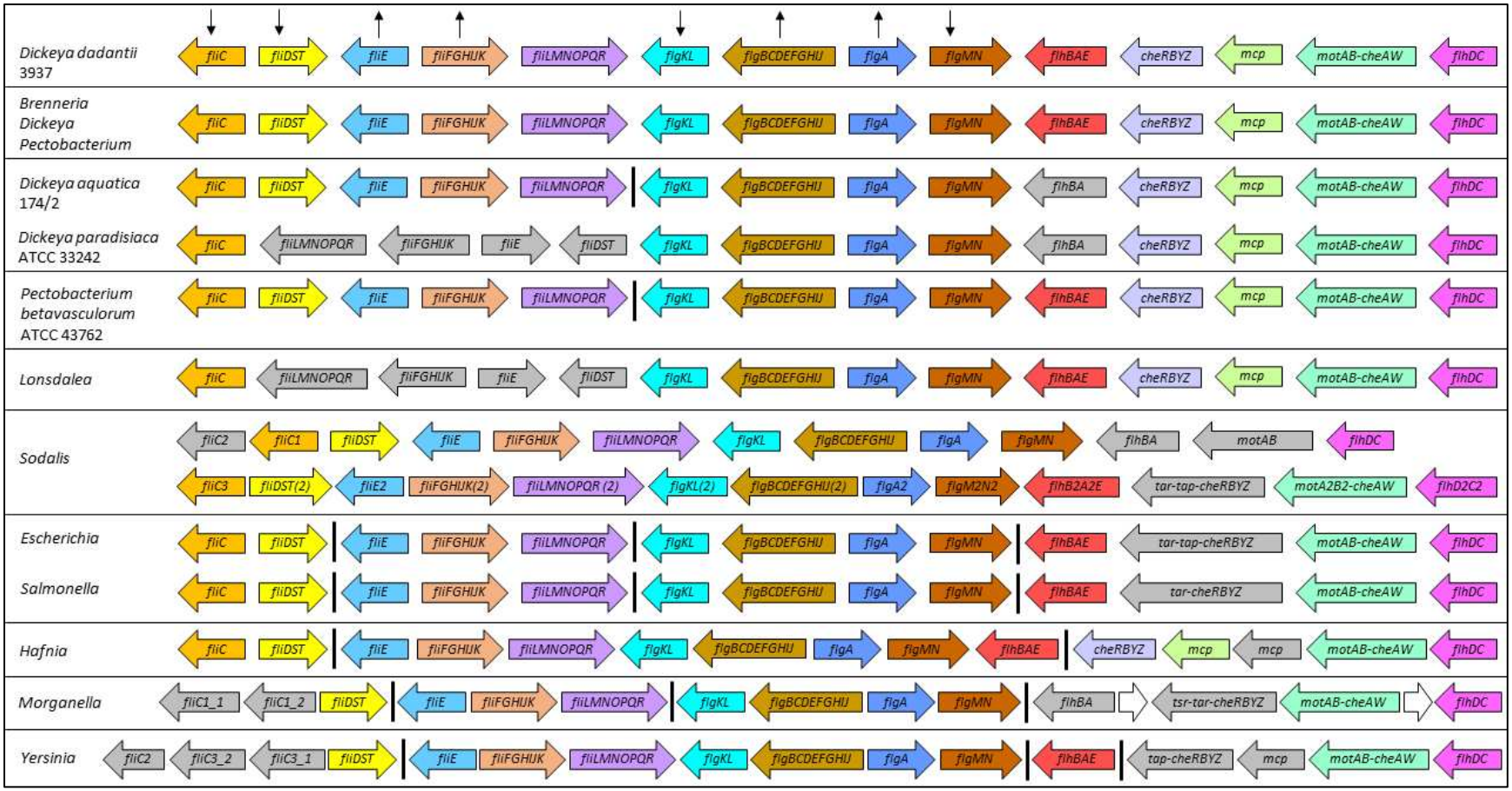

Figure 3.1: Conservation of flagellar gene organization and operon structure across the order Enterobacteriales. Arrowheads are indicative of genes and their direction. White arrowheads are hypothetical genes. Downward arrow represents gene/operon that was differentially down-regulated in *fliA* mutant compared to WT *D. dadantii* 3937. Upward arrow represents gene/operon that was differentially up-regulated in *fliA* mutant compared to WT *D. dadantii* 3937. Gray color illustrates the non-conserved flagellar gene organization. Vertical line indicates variable region containing non-conserved genes.

[1]The species and strain names were omitted for simplicity, except for four strains of *Dickeya* and *Pectobacterium*, which showed variation in flagellar gene organization and operon structure compared to other species in the same genera.

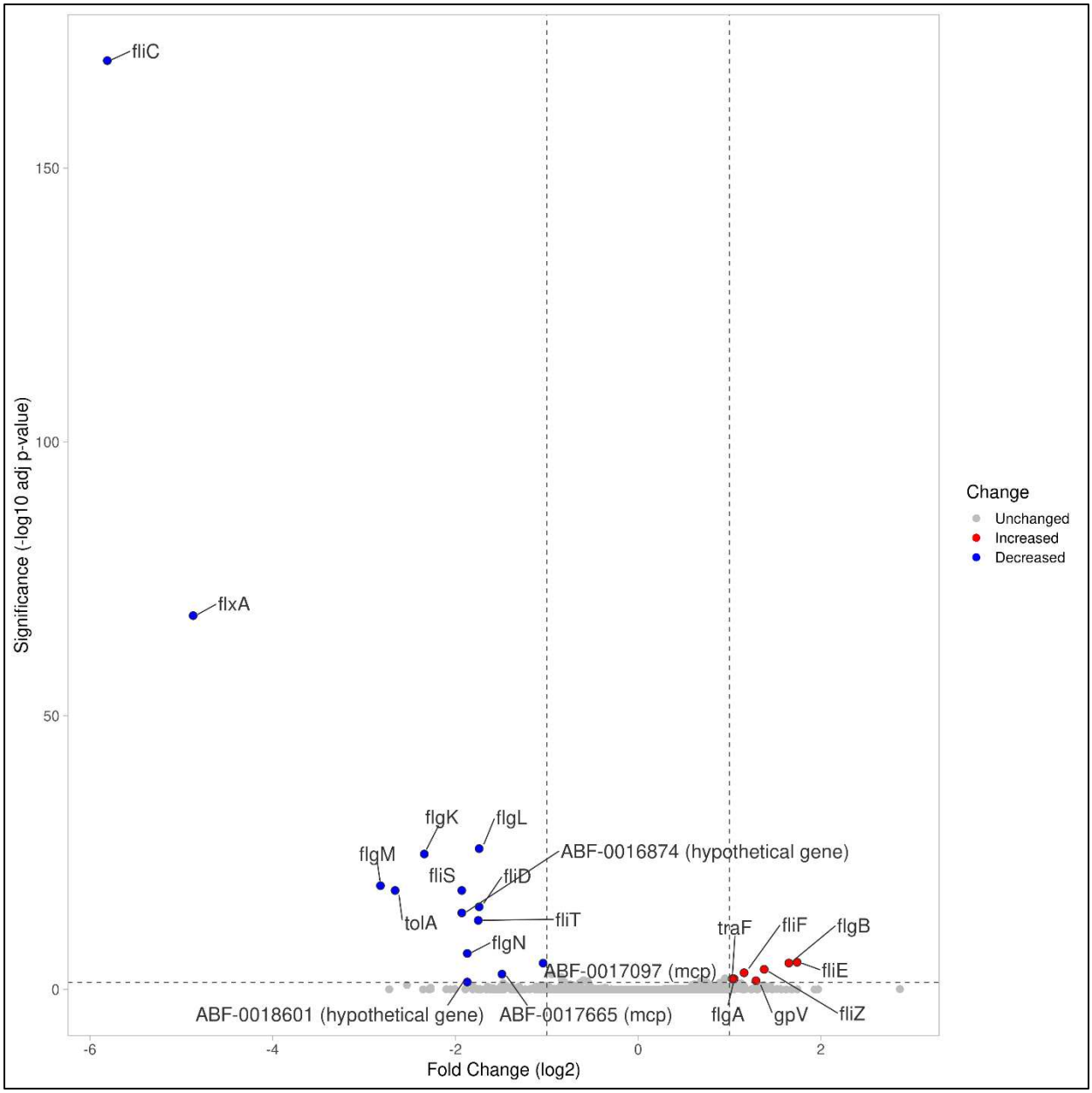

Figure 3.2: Volcano plot of expressed genes in *fliA* mutant compared to WT *D. dadantii* 3937 in MM-glu. The red dots denote significantly upregulated genes, the green dots represent signficantly downregulated genes (log2 fold change cut-off of 1 and -1; adjusted p-value $\leq$ 0.05), and the grey dots indicate non-significant expressed genes.

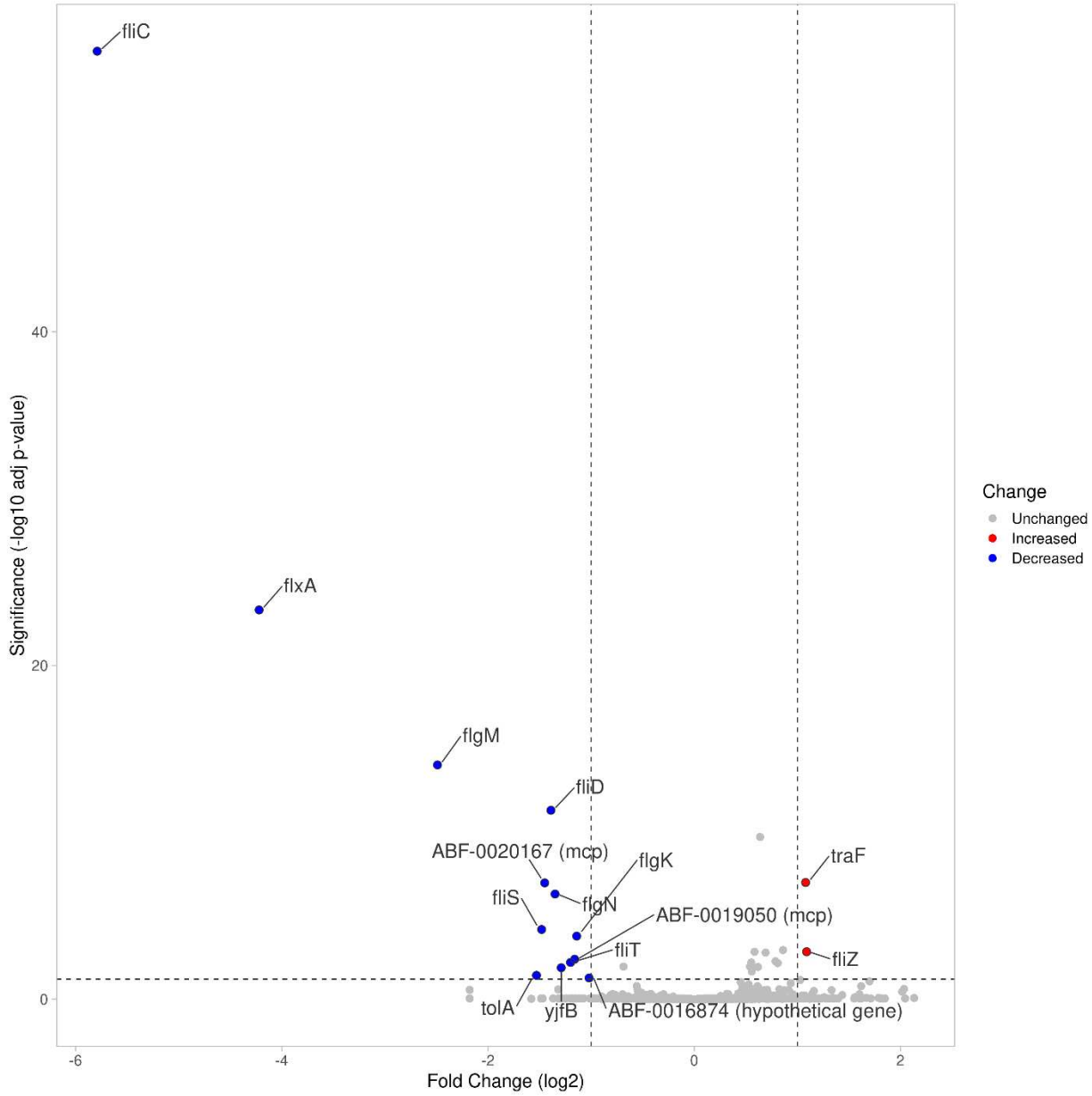

Figure 3.3: Volcano plot of expressed genes in *fliA* mutant compared to WT *D. dadantii* 3937 in MM-gly. The red dots denote significantly upregulated genes, the green dots represent signficantly downregulated genes (log2 fold change cut-off of 1 and -1; adjusted p-value $\leq$ 0.05), and the grey dots indicate non-significant expressed genes.

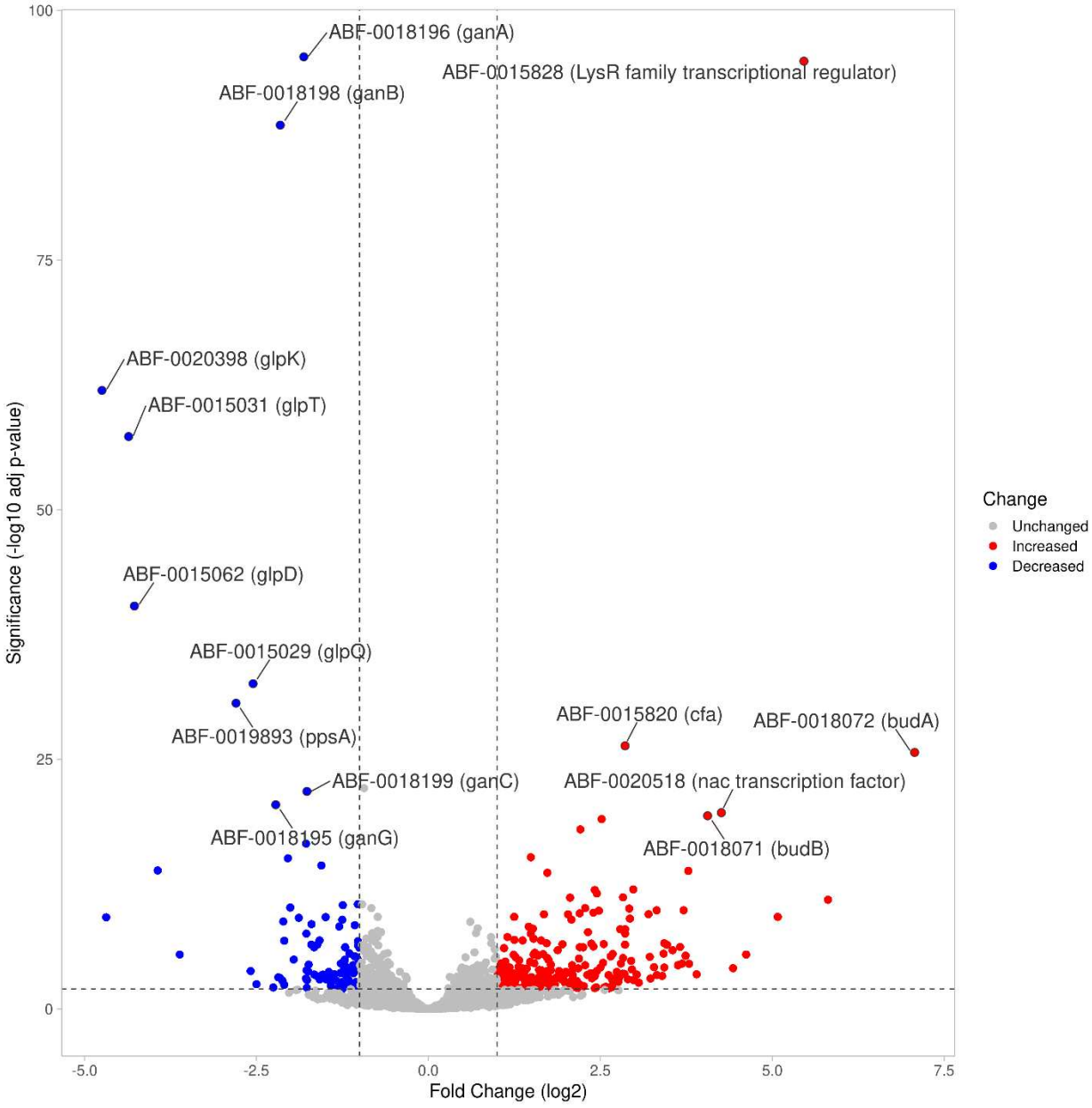

Figure 3.4: Volcano plot of expressed genes in WT *D. dadantii* 3937 in MM-glu vs MM-gly. The red dots denote significantly upregulated genes, the green dots represent signficantly downregulated genes (log2 fold change cut-off of 1 and -1; adjusted p-value $\leq$ 0.05), and the grey dots indicate non-significant expressed genes.

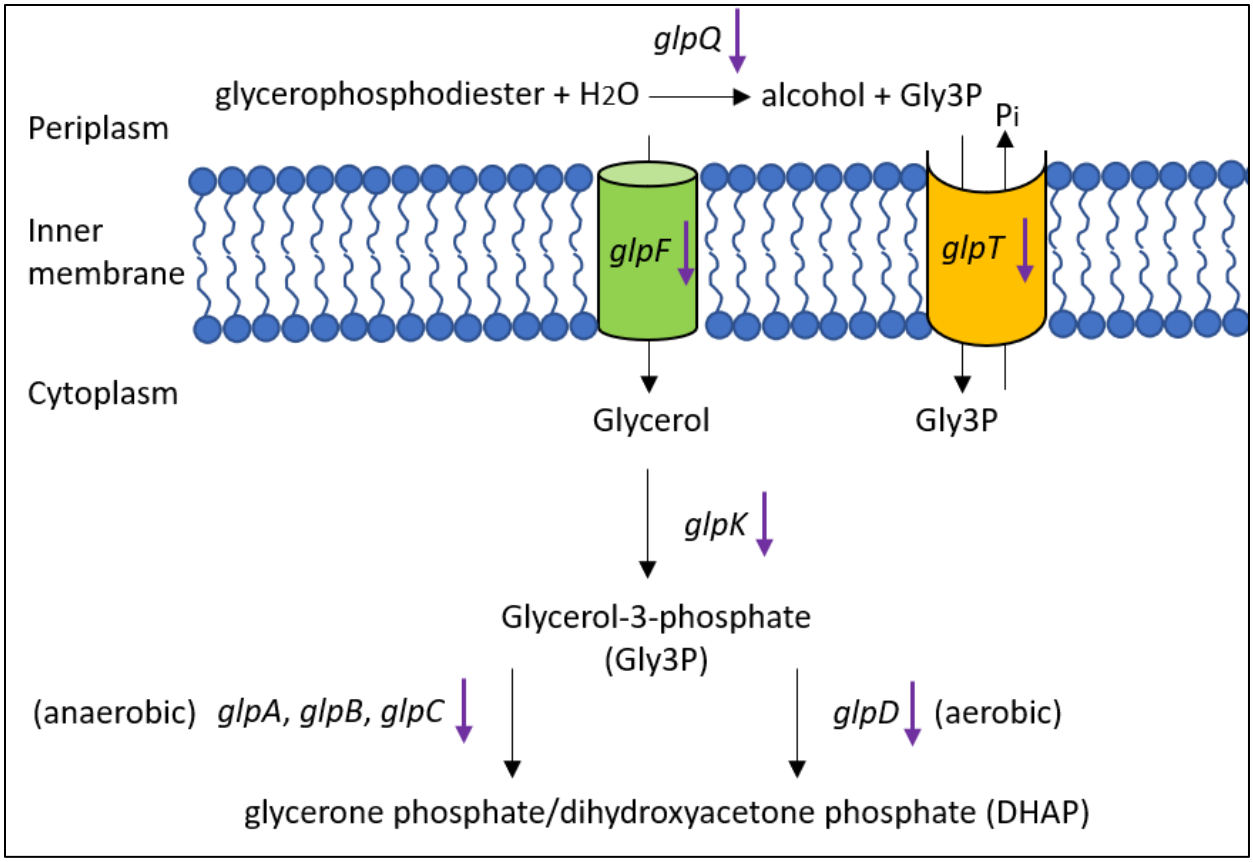

Figure 3.5: Genes involved in glycerol metabolism in *Dickeya dadantii* 3937 were differentially down-regulated in MM-glu compared to MM-gly. Purple arrow shows the downregulation of genes in MM-glu compared to MM-gly.

REFERENCES

1. Charkowski A, Blanco C, Condemine G, Expert D, Franza T, Hayes C, et al. The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. Annual Review of Phytopathology. 2012;50:425-49.

2. Charkowski AO. The changing face of bacterial soft-rot diseases. Annual Review of Phytopathology. 2018;56:269-88.

3. Río-Álvarez I, Muñoz-Gómez C, Navas-Vásquez M, Martínez-García PM, Antúnez-Lamas M, Rodríguez-Palenzuela P, et al. Role of *Dickeya dadantii* 3937 chemoreceptors in the entry to Arabidopsis leaves through wounds. Molecular Plant Pathology. 2015;16(7):685-98.

4. Yao J, Allen C. Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*. Journal of Bacteriology. 2006;188(10):3697-708.

5. Misas-Villamil JC, Kolodziejek I, van der Hoorn RA. *Pseudomonas syringae* colonizes distant tissues in *Nicotiana benthamiana* through xylem vessels. The Plant Journal. 2011;67(5):774-82.

6. Kumar Verma R, Samal B, Chatterjee S. *Xanthomonas oryzae* pv. *oryzae* chemotaxis components and chemoreceptor Mcp2 are involved in the sensing of constituents of xylem sap and contribute to the regulation of virulence-associated functions and entry into rice. Molecular Plant Pathology. 2018;19(11):2397-415.

7. Macnab RM. How bacteria assemble flagella. Annual Reviews in Microbiology. 2003;57(1):77-100.

8. Eisenbach M. Control of bacterial chemotaxis. Molecular Microbiology. 1996;20(5):903-10.

9. Ud-Din AIMS, Roujeinikova A. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. Cellular and Molecular Life Sciences. 2017;74(18):3293-303.

10. Parkinson JS, Hazelbauer GL, Falke JJ. Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. Trends in Microbiology. 2015;23(5):257-66.

11. Coleman MD, Bass RB, Mehan RS, Falke JJ. Conserved glycine residues in the cytoplasmic domain of the aspartate receptor play essential roles in kinase coupling and on− off switching. Biochemistry. 2005;44(21):7687-95.

12. Fitzgerald DM, Bonocora RP, Wade JT. Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. PLOS Genetics. 2014;10(10):e1004649.

13. Chilcott GS, Hughes KT. Coupling of Flagellar Gene Expression to Flagellar Assembly in *Salmonella enterica* Serovar Typhimurium and *Escherichia coli*. Microbiology and Molecular Biology Reviews. 2000;64(4):694-708.

14. Liu X, Matsumura P. Differential regulation of multiple overlapping promoters in flagellar class II operons in *Escherichia coli*. Molecular Microbiology. 1996;21(3):613-20.

15. Ide N, Ikebe T, Kutsukake K. Reevaluation of the promoter structure of the class 3 flagellar operons of *Escherichia coli* and *Salmonella*. Genes and Genetic Systems. 1999;74(3):113-6.

16. Dees MW, Lysøe E, Rossmann S, Perminow J, Brurberg MB. *Pectobacterium polaris* sp. nov., isolated from potato (*Solanum tuberosum*). International Journal of Systematic and Evolutionary Microbiology. 2017;67(12):5222-9.

17. Antúnez-Lamas M, Cabrera-Ordonez E, Lopez-Solanilla E, Raposo R, Trelles-Salazar O, Rodríguez-Moreno A, et al. Role of motility and chemotaxis in the pathogenesis of *Dickeya dadantii* 3937 (ex *Erwinia chrysanthemi* 3937). Microbiology. 2009;155(2):434-42.

18. Hugouvieux-Cotte-Pattat N, Blot N, Reverchon S. Identification of TogMNAB, an ABC transporter which mediates the uptake of pectic oligomers in *Erwinia chrysanthemi* 3937. Molecular Microbiology. 2001;41(5):1113-23.

19. Babujee L, Apodaca J, Balakrishnan V, Liss P, Kiley PJ, Charkowski AO, et al. Evolution of the metabolic and regulatory networks associated with oxygen availability in two phytopathogenic enterobacteria. BMC Genomics. 2012;13(1):110.

20. Hommais F, Oger-Desfeux C, Van Gijsegem F, Castang S, Ligori S, Expert D, et al. PecS is a global regulator of the symptomatic phase in the phytopathogenic bacterium *Erwinia chrysanthemi* 3937. Journal of Bacteriology. 2008;190(22):7508-22.

21. Zhao K, Liu M, Burgess RR. Adaptation in bacterial flagellar and motility systems: from regulon members to 'foraging'-like behavior in *E. coli*. Nucleic Acids Research. 2007;35(13):4441-52.

22. Delcenserie V, LaPointe G, Charaslertrangsi T, Rabalski A, Griffiths MW. Glucose decreases virulence gene expression of *Escherichia coli* O157: H7. Journal of Food Protection. 2012;75(4):748-52.

23. Yap M-N, Yang C-H, Barak JD, Jahn CE, Charkowski AO. The *Erwinia chrysanthemi* type III secretion system is required for multicellular behavior. Journal of Bacteriology. 2005;187(2):639-48.

24. Jahn CE, Willis DK, Charkowski AO. The flagellar sigma factor FliA is required for *Dickeya dadantii* virulence. Molecular Plant-Microbe Interactions. 2008;21(11):1431-42.

25. Miller J. Experiments in Molecular Genetics New York: Cold Spring Harbor Laboratory Press; 1972.

26. Jahn CE, Charkowski AO, Willis DK. Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. Journal of Microbiological Methods. 2008;75(2):318-24.

27. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.

28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.

29. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology. 2019;37(8):907-15.

30. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

31. Team RC. R: A language and environment for statistical computing. Vienna, Austria; 2013.

32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.

33. Goedhart J, Luijsterburg MS. VolcaNoseR is a web app for creating, exploring, labeling and sharing volcano plots. Scientific Reports. 2020;10(1):1-5.

34. Glasner JD, Liss P, Plunkett III G, Darling A, Prasad T, Rusch M, et al. ASAP, a systematic annotation package for community analysis of genomes. Nucleic Acids Research. 2003;31(1):147-51.

35. Glasner JD, Rusch M, Liss P, Plunkett III G, Cabot EL, Darling A, et al. ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Research. 2006;34(suppl_1):D41-D5.

36. Braun V, Mahren S, Ogierman M. Regulation of the FecI-type ECF sigma factor by transmembrane signalling. Current Opinion in Microbiology. 2003;6(2):173-80.

37. del Pilar Marquez-Villavicencio M, Weber B, Witherell RA, Willis DK, Charkowski AO. The 3-hydroxy-2-butanone pathway is required for *Pectobacterium carotovorum* pathogenesis. PLOS ONE. 2011;6(8):e22974.

38. Effantin G, Rivasseau C, Gromova M, Bligny R, Hugouvieux-Cotte-Pattat N. Massive production of butanediol during plant infection by phytopathogenic bacteria of the genera *Dickeya* and *Pectobacterium*. Molecular Microbiology. 2011;82(4):988-97.

39. Enard C, Diolez A, Expert D. Systemic virulence of *Erwinia chrysanthemi* 3937 requires a functional iron assimilation system. Journal of Bacteriology. 1988;170(6):2419-26.

40. Franza T, Expert D. The virulence-associated chrysobactin iron uptake system of *Erwinia chrysanthemi* 3937 involves an operon encoding transport and biosynthetic functions. Journal of Bacteriology. 1991;173(21):6874-81.

41. Grinter R, Josts I, Mosbahi K, Roszak AW, Cogdell RJ, Bonvin AM, et al. Structure of the bacterial plant-ferredoxin receptor FusA. Nature Communications. 2016;7(1):1-10.

42. Dahler G, Barras F, Keen N. Cloning of genes encoding extracellular metalloproteases from *Erwinia chrysanthemi* EC16. Journal of Bacteriology. 1990;172(10):5803-15.

43. Wandersman C, Delepelaire P, Letoffe S, Schwartz M. Characterization of *Erwinia chrysanthemi* extracellular proteases: cloning and expression of the protease genes in *Escherichia coli*. Journal of Bacteriology. 1987;169(11):5046-53.

44. Chatterjee A, Cui Y, Chatterjee AK. Regulation of *Erwinia carotovora hrpL*$_{Ecc}$ (*sigma-L*$_{Ecc}$), Which Encodes an Extracytoplasmic Function Subfamily of Sigma Factor Required for Expression of the HRP Regulon. Molecular Plant-Microbe Interactions. 2002;15(9):971-80.

45. Yap M-N, Rojas CM, Yang C-H, Charkowski AO. Harpin mediates cell aggregation in *Erwinia chrysanthemi* 3937. Journal of Bacteriology. 2006;188(6):2280-4.

46. Kazemi-Pour N, Condemine G, Hugouvieux-Cotte-Pattat N. The secretome of the plant pathogenic bacterium *Erwinia chrysanthemi*. Proteomics. 2004;4(10):3177-86.

47. Whitfield C. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. Annual Review of Biochemistry. 2006;75:39-68.

48. Narváez-Barragán DA, Tovar-Herrera OE, Segovia L, Serrano M, Martinez-Anaya C. Expansin-related proteins: biology, microbe–plant interactions and associated plant-defense responses. Microbiology. 2020;166(11):1007-18.

49. Rio-Alvarez I, Rodríguez-Herva JJ, Cuartas-Lanza R, Toth I, Pritchard L, Rodríguez-Palenzuela P, et al. Genome-wide analysis of the response of *Dickeya dadantii* 3937 to plant antimicrobial peptides. Molecular Plant-Microbe Interactions. 2012;25(4):523-33.

50. Koskiniemi S, Lamoureux JG, Nikolakakis KC, de Roodenbeke CtK, Kaplan MD, Low DA, et al. Rhs proteins from diverse bacteria mediate intercellular competition. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(17):7032-7.

51. Loth K, Costechareyre D, Effantin G, Rahbé Y, Condemine G, Landon C, et al. New Cyt-like δ-endotoxins from *Dickeya dadantii*: structure and aphicidal activity. Scientific Reports. 2015;5(1):1-10.

52. Shimada T, Fujita N, Yamamoto K, Ishihama A. Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. PLOS ONE. 2011;6(6):e20081.

53. Hénin J, Tajkhorshid E, Schulten K, Chipot C. Diffusion of glycerol through *Escherichia coli* aquaglyceroporin GlpF. Biophysical Journal. 2008;94(3):832-9.

54. Voegele RT, Sweet GD, Boos W. Glycerol kinase of *Escherichia coli* is activated by interaction with the glycerol facilitator. Journal of Bacteriology. 1993;175(4):1087-94.

55. Huang Y, Lemieux MJ, Song J, Auer M, Wang D-N. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. Science. 2003;301(5633):616-20.

56. Tommassen J, Eiglmeier K, Cole ST, Overduin P, Larson TJ, Boos W. Characterization of two genes, *glpQ* and *ugpQ*, encoding glycerophosphoryl diester phosphodiesterases of *Escherichia coli*. Molecular and General Genetics. 1991;226(1-2):321-7.

57. Larson T, Ehrmann M, Boos W. Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the glp regulon. Journal of Biological Chemistry. 1983;258(9):5428-32.

58. Iuchi S, Cole S, Lin E. Multiple regulatory elements for the *glpA* operon encoding anaerobic glycerol-3-phosphate dehydrogenase and the *glpD* operon encoding aerobic glycerol-3-phosphate dehydrogenase in *Escherichia coli*: further characterization of respiratory control. Journal of Bacteriology. 1990;172(1):179-84.

59. Green J, Stapleton MR, Smith LJ, Artymiuk PJ, Kahramanoglou C, Hunt DM, et al. Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches. Current Opinion in Microbiology. 2014;18:1-7.

60. Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). Journal of Molecular Biology. 1999;293(2):199-213.

61. Notley-McRobb L, Death A, Ferenci T. The relationship between external glucose concentration and cAMP levels inside *Escherichia coli*: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. Microbiology. 1997;143(6):1909-18.

62. Deutscher J. The mechanisms of carbon catabolite repression in bacteria. Current Opinion in Microbiology. 2008;11(2):87-93.

63. Weissenborn DL, Wittekindt N, Larson TJ. Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. Journal of Biological Chemistry. 1992;267(9):6122-31.

64. Pédron J, Chapelle E, Alunni B, Van Gijsegem F. Transcriptome analysis of the *Dickeya dadantii* PecS regulon during the early stages of interaction with *Arabidopsis thaliana*. Molecular Plant Pathology. 2018;19(3):647-63.

65. Yang S, Peng Q, Zhang Q, Zou L, Li Y, Robert C, et al. Genome-wide identification of HrpL-regulated genes in the necrotrophic phytopathogen *Dickeya dadantii* 3937. PLOS ONE. 2010;5(10):e13472.

66. Dong T, Yu R, Schellhorn H. Antagonistic regulation of motility and transcriptome expression by RpoN and RpoS in *Escherichia coli*. Molecular Microbiology. 2011;79(2):375-86.

67. Chanda B, Xia Y, Mandal MK, Yu K, Sekine KT, Gao Q-m, et al. Glycerol-3-phosphate is a critical mobile inducer of systemic immunity in plants. Nature Genetics. 2011;43(5):421-7.

68. Hu J, Zhang Y, Wang J, Zhou Y. Glycerol affects root development through regulation of multiple pathways in Arabidopsis. PLOS ONE. 2014;9(1):e86269.

69. Yang Y, Zhao J, Liu P, Xing H, Li C, Wei G, et al. Glycerol-3-phosphate metabolism in wheat contributes to systemic acquired resistance against *Puccinia striiformis* f. sp. *tritici*. PLOS ONE. 2013;8(11):e81756.

70. Barber AM, Zhurkin VB, Adhya S. CRP-binding sites: evidence for two structural classes with 6-bp and 8-bp spacers. Gene. 1993;130(1):1-8.

Chapter 4

Future Directions

1. What are the specific ligands for individual methyl-accepting chemotaxis proteins?

   In Chapter 1, preliminary studies showed that soft rot Pectobacteriaceae were attracted to ribose, aspartic acid, and malic acid. Efforts to determine the specific ligands for methyl-accepting chemotaxis proteins (MCP) in *D. dianthicola* failed due to growth defects in the transformants. An alternative way to determine this is via isothermal titration calorimetry (ITC) assay using only the ligand binding domain of individual MCP (1). The ligand binding domain of MCP can be cloned into a plasmid and overexpressed in *E. coli* (1). The protein of the ligand binding domain of MCP can then be purified and subjected to ITC assay (1). ITC is a quantitative method that determines the binding affinity between a protein or DNA to its ligand by measuring the amount of heat released or absorbed during the reaction.

2. How are chemotaxis genes regulated in *Dickeya*?

   In Chapter 3, we found that FliA does not regulate methyl-accepting chemotaxis genes (*mcp*) in *D. dadantii* in minimal media, which indicates that other sigma factors are involved in *mcp* transcription initiation. A potential candidate is the nitrogen-limitation sigma factor, RpoN. RNA-sequencing experiment can be carried out using the wild-type strain of *D. dadantii* 3937 grown in nitrogen-rich medium and nitrogen-limited medium. If RpoN regulates *mcp* expression, suppression of chemotaxis could serve as a potential management of soft rot disease of potato through adequate nitrogen fertilization of plants.

3. Future management of soft rot disease of potato

   Based on the preliminary results of the chemical attractants for soft rot Pectobacteriaceae in Chapter 1, future management of soft rot disease of potato could be done using a slow-release chemoattractant in the soil to attract the bacteria away from host plants.

REFERENCES

1. Cerna-Vargas JP, Santamaría-Hernando S, Matilla MA, Rodríguez-Herva JJ, Daddaoua A, Rodríguez-Palenzuela P, et al. Chemoperception of specific amino acids controls phytopathogenicity in *Pseudomonas syringae* pv. tomato. mBio. 2019;10(5).